ED n^{o} 432 : «Sciences des Métiers de l'Ingénieur »

N^{o} attribué par la bibliothèque
|_|_|_|_|_|_|_|_|_|_|

T H E S E

pour obtenir le grade de :
DOCTEUR DE L'ECOLE DES MINES DE PARIS
Spécialité "Energétique"

présentée et soutenue publiquement par
Nils SIEBERT

le 06 mars 2008

DEVELOPMENT OF METHODS FOR REGIONAL WIND POWER FORECASTING

(DÉVELOPPEMENT DE MÉTHODES POUR LA PRÉDICTION DE LA PRODUCTION ÉOLIENNE RÉGIONALE)

Directeur de thèse: Georges KARINIOTAKIS

Jury :

Pr. Didier MAYER .... Président
Pr. Henrik MADSEN .... Rapporteur
Dr. Eric PEIRANO .... Rapporteur
Dr. Ismael SÁNCHEZ .... Rapporteur
Pr. Xavier GUILLAUD .... Examinateur
Dr. Georges KARINIOTAKIS .... Examinateur

Acknowledgments

In 2003 I was searching for a subject for my master's Thesis. Through no fault of my own, I stumbled on an offer from the then Centre d'Energétique in Sophia Antipolis, who was looking for a student to conduct research on the adaptation of the F-NN model to the case of regional wind power forecasting. I applied and got the position. Five years latter I have the privilege of defending a Phd Thesis on the same subject. For this, I would like to thank Dr. George Kariniotakis and Prof. Didier Mayer for the opportunity they gave me. I would also like to thank the ADEME, the French Agency for Environment and Energy Management, for funding my research.

Writing thesis would not been possible without the help of my supervisor, Dr. George Karinotakis. His advice and in-depth review of my work were crucial to its successful completion.

I would also like to thank all the members of the jury for accepting to judge my work (and finding it worthy). Special thanks go to Prof. Henrik Madsen, Dr. Ismael Sanchez and Dr. Eric Peirano, who found time in their busy schedules to referee the manuscript. Their comments and suggestions are priceless for the continuation of my work.

Professional acknowledgments

I would like to thank all the partners of the European project Anemos. Particular thanks go to ESB National Grid for the data on Irish wind farms and Elsam (now Dong) for the data on Danish production. Thanks also to all the people involved for making it an enjoyable experience: exploring Crete with Gaël and Pierre is memorable, as well as late night discussions with Torben, Gregor and Vincent in Athens' Plaka district. More seriously, the discussions and scientific emulation were very fruitful and a large part of the work presented in this thesis stems from this collaboration.

Many thanks go to Dr. Pierre Pinson who supervised my Master's thesis, introducing me to the wonderful world of wind power forecasting, and with whom I latter collaborated during the early years of my research work. Pierre continues to work in the field and I very much look forward to continuing our collaboration.

Thanks also to Prof. Lucien Wald for his eye opening discourse on the first day of the Phd Course, and later for his constructive criticism at the yearly evaluation seminars. Lucien, I will keep working on my oral presentation skills. In the same vein, I would like to thank Dr. Thierry Ranchin for refereeing my yearly evaluation reports. His comments and suggestions helped me stay on track.

I would also like to thank all my colleagues from the Center of Energy and Processes for creating an enjoyable and stimulating environment. Special thanks go to the IT team, Patrick, Alain and Laurent; thank you for putting up with my frequent requests and for your self-control when I accidentally crashed some servers.

Finally I would like to thank all the teachers I have had for the knowledge they passed on. Special thanks go to all the faculty at the Institut de Mathématiques Appliquées of the Université Catholique de l'Ouest, and I would like to particularly thank Pierre Chauvet, Philippe Pons and Jean Magat, for transmitting their love of mathematics and computer science.

Personal acknowledgments

I would like to thank my parents. Thank you for love, your support, your guidance and all the wonderful things you have given me. I hope I will be as good a parent as you are. This thesis is dedicated to you.

I would also like to thank my friends and colleagues Alexis, Jeremie and Luis. Alexis, working with you has thought me a lot. Jeremie and Luis, thank you for putting up with my presence in the same office through all these years. Thank you also for all the stimulating conversations and debates, they inspired many of the ideas presented in this work. Hang in there guys, you will soon finish your theses.

Last but not least, Alexandra, who found the master's thesis proposal on the web and who has been the fount of inspiration for my work. Thank you for your love and your support through the good times and the bad. I will never thank you enough for putting up with me during the last five years and for wanting to put up with me for the rest of your life.

Contents

List of Figures

List of Tables

Abbreviations

a.g.l.	Above ground level
AI	Artificial Intelligence
APX	Amsterdam Power eXchange
ARMA	Auto-Regressive Moving Average
ARMAX	Auto-Regressive Moving Average with exogenous variables
ARXM	Auto-Regressive with exogenous variables
AWPPS	Armines Wind Power Prediction System
CENER	Centro Nacional de Energías Renovables
CEP	Center For Energy and Processes
CFD	Computational Fluid Dynamics
CH_4	Methane
CIEMAT	Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas
CO_2	Carbon Dioxide
CPU	Central Processing Unit
DWD	Deutscher Wetterdienst
ECMWF	European Centre for Medium-range Weather Forecasts
EU	European Union
EWEA	European Wind Energy Association
F-NN	Fuzzy-Neural Networks
GWEC	Global Wind Energy Council
HIRLAM	High Resolution Limited Area Model
IEA	International Energy Agency
IMM	Informatics and Mathematical Modeling department of the Technical University of Denmark
ISET	Institut für Solare Energieversorgungstechnik
MAE	Mean Absolute Error
MS-EPS	Multi-Scheme Ensemble Prediction System
MIFS	Mutual Information Based Feature Selection
MOS	Model Output Statistics
MSE	Mean Square Error
NCEP	National Center for Environmental Prediction
NMAE	Normalized Mean Absolute Error
NRMSE	Normalized Root Mean Square Error
NSDE	Normalized Standard Deviation of the Errors
NWP	Numerical Weather Prediction
PCA	Principal Component Analysis
RAMS	Regional Atmospheric Modeling System

RES	Renewable Energy Source
RMSE	Root Mean Square Error
RPC	Regressive Power Curve
SCADA	Supervisory Control And Data Acquisition
SDE	Standard Deviation of the Errors
SOWIE	Simulation Model for the Operational Forecast of Wind Energy Production in Europe
SVM	Support Vector Machine
TSO	Transmission System Operator
WEPROG	Weather & wind Energy Prognosis
WPMS	Wind Power Management System
WPPT	Wind Power Prediction Tool

Notations and Mathematical Symbols

$\tilde{*}$	T-norm operator
$\mathcal{A}$	A set of wind farms
A	A fuzzy set
A_r	Rotor swept area
c_k	Sample autocovariance
c_p	Turbine performance coefficient
cov	Covariance
C_k^n	Number of possible combinations of k elements chosen among n
$d_{i,j}$	Distance between wind farms i and j
D	Divergence of two probability distributions
ϵ	Normalized prediction error
e	Prediction error
E	Expected value
f^{ar}	Sub-area forecasting model
f^{reg}	Regional forecasting model
f^{wf}	Single wind farm forecasting model
f_X	Probability density function of random variable X
η	Learning rate of the F-NN model
H	Information entropy of a random variable
I	Average mutual information between random variables
κ	Kurtosis of a probability distribution
k	Prediction horizon
L	Loss function
μ	Mean of a probability distribution
$M_d(t)$	Moving average of order d
ν	Skewness of a probability distribution
NMAE_{Sub}	NMAE performance criterion associated to a variable subset
NRMSE_{Sub}	NRMSE performance criterion associated to a variable subset
ψ	Membership function associated to a fuzzy set
p_X	Probability mass function of random variable X
P	Wind power
$\overline{P(t)}$	Average of all power measurements available at time t
P^{reg}	Total power output of a region
P^{wf}	Power output of a wind farm
P_{inst}	Installed capacity

P_X	Probability distribution of random variable X
P_{inst}^{wf}	Installed capacity of a wind farm
P_{inst}^{reg}	Total installed capacity of a region
ρ_{air}	Air density
$\rho_{X,Y}$	Correlation coefficient between random variables X and Y
r	Rotor diameter
r_k	Autocorrelation
r_{xy}	Pearson's correlation coefficient
R	Risk functional
R^2	Coefficient of determination
σ	Standard deviation of a probability distribution
$\sigma_{ensemble}$	NSDE of the forecast error of the regional production
σ_{single}	Mean NSDE of the forecast error computed at individual wind farms
s_x	Sample standard deviations of x
se_{r_k}	Standard error of the autocorrelation r_k
S	Smoothing factor of regional production
$Stab$	Transition stability between subsets
t	Time
v	Wind speed at hub height
V	Variance
$V_d(t)$	Moving variance of order d
Wd	Wind direction
Ws	Wind speed
wf_i	Wind farm i
$\overline{x}$	Sample mean of x

CHAPTER 1

Introduction

1.1 General context

RECENT studies have shown that over the last century the Earth's average surface temperature[1] has increased significantly. A report published by the International Panel on Climate Change states that the surface temperature increased by $0.6 \pm 0.2°C$ over the 20th century [1]. The same source also reports that the present atmospheric concentration of carbon dioxide (CO_2) and methane (CH_4) has not been exceeded in the past 420,000 years, and that these gases' concentrations have increased by 31% and 151% respectively since 1750. In this and other studies, it has been suggested that the growing concentrations of greenhouse gases in the atmosphere have greatly contributed to the observed global warming. Furthermore, since this increase in concentration in greenhouse gases closely follows the development of fossil-fuel use in the past 150 years, it has been suggested that anthropogenic greenhouse-gas emissions may, in part, be responsible for the current observed warming. These findings have led public opinion to push for carbon emission reduction measures in many countries.

Another recent development has been the realization that the current fossil-fuel based energy system is not sustainable and encompasses several risks. Today fossil-fuels dominate the global energy mix. Oil is primarily used in the transport sector whereas natural gas and coal are mainly used for electricity generation. In the future, the International Energy Agency (IEA) foresees a 60% increase in energy demand from now to 2030 with a doubling

[1] The average of near surface air temperature over land, and sea surface temperature

of electric demand [2]. Although the IEA does not foresee a peak in the production of conventional oil over the next three decades, increasing demand in fossil fuels from developing countries, and to a lesser extent from developed ones, will strain production capacity. Fossil-fuel importing countries will become increasingly dependent on imports from a diminishing, often politically unstable, number of suppliers. This reduction in the number of suppliers will also exacerbate the risk of supply disruption at strategic bottlenecks through which oil and gas are delivered.

Increasing environmental concerns over global warming and conventional energy resource depletion have led governing bodies to look for cleaner (in the sense of CO_2 emissions) and sustainable energy sources. Given the expected growth of electricity demand, particular attention has been given to the reduction of electricity consumption, to the development of more efficient conventional power generation systems and to the use renewable energy sources.

In many countries, policies designed to encourage the use of efficient energy conversion systems and to develop renewable energy sources have been implemented [3]. With this objective, the European Union adopted in 2001 the Directive on the promotion of electricity produced from renewable energy sources. This directive establishes targets for Renewable Energy Sources (RES) use for electricity production purposes for all member states. In particular, the directive states that 12% of gross national energy consumption, and 22.1% of electricity production should be met by RES by 2010[4]. In the United States of America, more than fifteen States have established policies designed to promote the development of renewable energy sources [5]. These are often minimum renewable energy requirements for utilities termed Renewable Portfolio Standards. Another example is India, where the use of all major renewable energy sources (solar, wind, small hydro and biomass) is being encouraged through different support schemes [6].

Most RES can be used to generate electricity. The most common technologies today are: hydro, solar photovoltaic, concentration solar, wind, biomass, and geothermal. Although most of these technologies have attained a certain maturity, wind power is witnessing the fastest development. The global wind power capacity has grown from 4800 MW in 1995 to more than 74223 MW in 2006 [7]. The Global Wind Energy Council (GWEC) predicts that by 2010, 149500 MW could be installed worldwide. This important development can be linked to several factors.

The first factor is the availability of the resource, there is wind everywhere on the planet. In a study by Hoogwijk [8], the global onshore economic potential of wind energy, defined as the amount of energy that can be technically produced given the cost of alternative energy sources, was evaluated. The results of this study showed that an amount equal to the 2001 global electricity consumption could be generated for less than 0.07 US$/kWh. When looking at the ratio between the technical potential and the 1996 electricity consumption at regional level, of the 17 regions defined in this study, only two regions (South East Asia and Japan) have a ratio lower than one. Another factor favouring the use of wind energy over other RES is its cost. Wind power is one of the most competitive technologies [2]. Although some other technologies, such as hydropower and geothermal are cheaper, they are more geographically constrained. In addition to the above two factors, wind energy has been given important support by policymakers. Effective support schemes have been implemented in many countries, the results of which can be seen in the evolution of the global wind energy market. In 2001, only five countries had reached the 1000 MW mark of installed capacity; four years later, eleven countries had reached this mark [9].

This large-scale development has led to an increase of wind power penetration in many power systems[2]. In some systems, wind penetration, in terms of energy, can reach levels as high as 15% to 20% [10]. Such high penetration levels can translate to power penetration levels of up to 100% a few hours a year, as is the case in the western Danish network. This can be problematic for power system operators because, unlike conventional power sources, wind power is variable and non-dispatchable. System operators cannot rely on wind power to maintain system balance. To integrate large amounts of wind power in electricity networks, the traditional solution is to schedule extra reserves in order to be able to meet unexpected wind power variations. In many cases this extra reserve allocation is provided by conventional fossil-fuel based plants. Consequently, it induces added system operating costs as well as pollutant emissions, thus reducing the advantages expected of wind power. One possible solution to this problem could be the use of energy storage devices to smooth wind power variations. Studies [11–14] have been conducted to analyse the possibility of palliating wind power variability using energy storage systems or through the use of other dispatchable RES, such as hydropower, and indicate that such a combination can be economically viable especially in the liberalized electricity markets. However, to our knowledge, no such explicitly combined power plants have been implemented on a large scale. Indeed, investment costs for storage technologies for capacities comparable to installed wind capacities can be prohibitively high.

To maintain system balance, utilities and TSOs have to plan the generation means necessary to cover the forecast load. Tools such as load forecasting, unit commitment and economic dispatch algorithms are used to optimize production. Because of its variability and non-dispatchable nature, wind power is traditionally considered as negative load [15–18]. As such, one possibility to alleviate some of the problems posed by wind power is to use wind power forecasts in the same way that load forecasts are used to help establish production schedules and minimize operational costs. In fact, wind power forecasts are considered as necessary, even mandatory [19], in systems where penetration is high, or in which little reserve capacity is available. The need for accurate wind power forecasting is recognized today by the power system industry as a crucial research field that needs to be addressed to allow efficient large-scale grid integration of wind power [20, 21].

[2]Wind penetration is defined as a ratio. Differences exist in the literature with respect to the values considered for this ratio. This can be the ratio between installed wind power capacity and the power system's total installed capacity, or the ratio between the energy provided by wind and the total energy demand of the system over a certain period, or finally as a power ratio between the amount of power fed into the system at a given time and the load at that time.

1.2 Wind power forecasting

Knowing the future has been one of Man's dearest wishes since the dawn of time. This desire has been one of the driving forces behind scientific research in many fields. For centuries, astronomical models have been developed to predict the position of planets. In other areas, physical laws have been defined to describe the outcome of different processes. With the advent of widespread, high-speed computing, forecasting capabilities have been extended to fields such as meteorology where calculation capacity was a major constraint [22]. As a result, it is now possible to predict complex processes such as wind power production with reasonably good accuracy.

Wind power forecasting aims to provide end-users with estimates of the likely available wind power at a given time in the future. Several prediction horizons can be considered. The case of interest here is short term forecasting where forecasts for the next 48 to 120 hours ahead with an hourly resolution are to be provided. This forecast horizon and resolution are considered because they are traditionally used for unit commitment purposes [23] or for participation in the electricity markets. The most common output of wind power forecasting models are spot forecasts; for each time step in the future, a single power value is provided. However, probabilistic forecasting models are being developed which provide additional information on the expected distribution of wind power.

To derive a short-term wind power forecast it is necessary to couple the output of a Numerical Weather Prediction (NWP) model and a conversion model. The conversion model must be capable of determining the expected wind farm production using as input NWP output and on-site measurements of production and/or meteorological variables. The conversion models are usually referred to as wind-power prediction models. Different approaches exist to compute the conversion. However, from a general perspective, these models can be seen as functions that model the output of a wind park based on the expected meteorological situation. From this point of view, having a good understanding of the phenomena conditioning the production of a group of wind turbines is necessary in order to establish models that can accurately approximate their behaviour.

Most models developed so far are mainly intended for forecasting the expected output of wind farms that are often considered as single power plants. As the number of wind farms increases, end-users express the need for forecasts of the expected aggregated output of all farms within a specified geographical region. In the case of power system operation, forecasts of the expected aggregated power output of all farms installed within a region are needed for various functions such as congestion management, reserves estimation, planning exchanges with neighbouring systems etc. Regional wind power forecasting is thus defined as the prediction of the aggregated power output of wind farms spread over a geographical region.

1.3 Regional wind power forecasting

Regional wind power forecasting deals with the prediction of the aggregated power output of wind farms located within a defined region. Regional wind power prediction differs from single wind farm forecasting in several aspects. The first one is that regional wind power production does not have the same characteristics as single wind farm production. A second aspect, is that the data available to regional prediction models is not necessarily as complete as that available to single wind farm models. Because of these specificities, specialized models are needed to provide the best possible forecasts of regional wind generation.

The main characteristic of regional production is that the considered installed capacity is dispersed over a wide geographical area. This dispersion has one main advantage, which is the smoothing of the aggregated output of the wind farms. By summing the output of several wind farms, individual power variations are compensated and the resulting production presents much slower variations. This statistical smoothing effect leads to the main characteristic of regional forecasting, which is that geographically dispersed production can be more accurately predicted than the production of single wind farms [24].

At a regional or national scale, the ideal case would be to predict the output of each single wind farm in order to obtain the total generation in the next hours. However, wind power forecasting models require input data concerning the wind farms to compute the forecasts. This information can be of two natures, static and dynamic. Static information is the information necessary to describe a wind farm: the rated power, the power curve of the turbines, etc. This information is generally considered as not evolving over time. Dynamic information is time series data that evolves over time such as production measurements and NWP forecasts for the wind farm.

In power systems with an important number of installed wind farms, predicting the output of each single farm may be impossible due to a lack of adequate descriptions of the wind farms (static data) and/or production measurements (dynamic data). For instance, on-line information is usually not available for all wind farms. This is because Supervisory Control and Data Acquisition (SCADA) systems may not be installed in all farms. It can be noted that grid codes in some countries impose such infrastructure for wind farms whose installed capacity is above a certain level.

The absence of SCADA systems means that the actual total wind power production is seldom known accurately. However, even if the actual power each wind park injects into the grid at a given time can be known with some delay, the actual available capacity at that time in the region is difficult to obtain. Here the available capacity refers to sum of the capacities of the turbines in operation at a given time. This can be particularly hampering for statistical model training since the measured regional power cannot be corrected for unavailable capacity, which leads to noise in the training data. This problem is similar to that encountered in forecasting situations where planned maintenance information is unavailable to the forecasting model.

Because of the usual lack of static and especially dynamic information, the mainstream approach to regional prediction is to use upscaling methods. The idea behind these methods is to use the available static and dynamic information to extrapolate the best possible forecast of the aggregated power output of all the farms in an area. In this work, the wind farms for which dynamic data is available will be referred to as reference wind farms.

1.4 Purpose of the thesis

The purpose of the work presented in this thesis is to develop a framework and guidelines that can be useful to modellers wishing to implement regional wind power forecasting models. More specifically, the purpose is to provide statistical model developers with answers to the following questions: what modelling strategies exist and what is their influence on forecast accuracy? How does the choice of statistical model influence forecast accuracy? How does input data influence forecasting model performance and how can selection of explanatory variables be performed?

To provide answers to these questions we first propose a framework for the characterization of the regional wind power production from a time series point of view. Regional wind power production is analyzed both as a single time series and as the aggregate of several wind farm productions. Then we investigate the relation between regional wind power production and NWP forecasts. In this way, salient aspects of the regional wind power forecasting problem, which must be taken into account when designing a regional forecasting model, are identified. This characterization also provides results that allow a more detailed assessment of the performance of regional wind power forecasting models.

We then examine the regional forecasting problem from a statistical learning perspective. Following a review of state-of-the-art regional forecasting solutions, we first define three basic approaches that can be used to combine models (usually single wind farm models) to build regional models. We then concentrate on examining the influence of different combination schemes on forecast accuracy. The aim is to determine if one combination scheme is superior to others. From this we proceed to investigate the influence that the choice of sub-model can have on the performance of a regional forecasting model combination. To achieve this we compare the performance of two different models used in different combination schemes. This comparison is made possible by the introduction of a novel forecasting model whose performance is shown to be comparable to that of other state-of-the-art models found in the literature. From these two examinations, conclusions are provided on the relative influence of model combination and of model choice, on regional wind power forecasting accuracy.

Although the type of model and the combination approach impact forecast accuracy, an important source of model accuracy, or inaccuracy, are the explanatory variables used to model the process. To examine the impact of explanatory variables on forecast accuracy, we first provide a description of the two types of variables used in statistical wind power forecasting: measured and forecast variables. The influence of these two types is examined in the frame of regional wind power forecasting, in which the number of potentially available variables exposes the forecasting models to the curse of dimensionality. More specifically, the forecast accuracy obtained with all the combinations of the available variables are examined. From this analysis, general guidelines are derived with respect to the number of variables to be considered and the usefulness of each variable type. Further, given the important number of relevant and often redundant explanatory variables available in the frame of regional wind power forecasting, explanatory variable selection techniques are investigated. Two variable selection methods, based on the exploitation of problem specific characteristics, are proposed. These are compared to a state-of-the-art method and the performance of all three is investigated in detail. Together, these methods constitute a first answer to the problem of input variable selection in the frame of regional wind power forecasting.

1.5 Structure of the thesis

In order to introduce the reader to the problem of wind power forecasting chapter 2 describes the state of the art in wind power forecasting. In this literature review the basis of wind power forecasting is presented. An overview of existing wind power forecasting models is presented along with the criteria commonly used to evaluate forecasting model performance. Recent developments in the areas of wind power forecast uncertainty estimation and the economic value of wind power forecasts are then reported. Our survey then concentrates on regional wind power forecasting models that have been proposed in the literature. From this review, lacks in the regional wind power forecasting literature and current research are identified and the objectives of the thesis justified.

The aim of chapter 3 is to characterize the regional wind power production and its relation to available explanatory variables. This characterization will serve as the basis of the later developments in the thesis. The chapter begins by formalizing the regional wind power forecasting problem. A framework is then proposed to examine regional wind power production and its statistical relation to available explanatory variables, namely NWP forecasts. This framework considers regional wind power production both as a single time series and as the aggregate of multiple time series and allows examining its properties from both perspectives. The relation between regional production and available NWP variables is characterized using a measure of the mutual information that exists between the different variables. This characterization framework is then applied to two case studies: one comprising 11 wind farms in Ireland and one considering the total regional production of the Jutland peninsula and of Funen island in Denmark. The results of the characterization of both cases are analyzed and compared to derive general conclusions on the most salient features of the regional wind power forecasting problem.

In chapter 4 the regional wind power forecasting problem is approached from a statistical learning perspective. From this perspective, existing regional wind power forecasting models are examined and three basic base-line model (or sub-model) combination schemes are identified. The work then concentrates on characterizing the performance of different base-line models used in these three approaches. Two base-line models are used in this comparison: the state-of-the-art Fuzzy-Neural Network model developed by

Kariniotakis [25] and a model developed in this thesis, named Regressive Power Curve (RPC) model. The performance of the RPC model is first validated by comparison to the performance of other state-of-the-art models on three case studies corresponding to different terrain complexities. The RPC model is shown to have a forecast accuracy comparable to that of other state-of-the-art models. The two base-line models are then used to investigate the relative strengths of the different regional wind power forecasting approaches and to examine the influence of using different base-line models in these approaches. The evaluation is performed on the Irish and Danish case studies described above.

The general problem of the impact of explanatory variables on forecast accuracy is examined in chapter 5. In the first part of this chapter a methodology is proposed to examine the influence of different types of explanatory variables on regional wind power forecast accuracy. This methodology is consequently applied explanatory variables such as measured production time series and wind speed forecasts. From the detailed analysis of the results, general conclusions are derived and the necessity of performing variable selection is clearly identified. This leads to the second part of the chapter, which investigates the application of explanatory variable selection techniques to the regional wind power forecasting problem. This part starts with an overview of existing variable selection methodologies. From this review, three methods, based on different methodologies, are proposed. One is a state-of-the-art filter method taken as an "off the shelf" solution. The two other methods, a clustering-based filter and a stability-based wrapper, are proposed here. They are adaptations of existing methodologies and exploit certain characteristics of the regional wind power forecasting problem. The three methods are benchmarked on the Irish and Danish case studies. From this benchmark, a detailed analysis of the properties of the methods is carried out and general conclusions on their usefulness are provided.

Finally, chapter 6 provides general conclusions and guidelines on the topics investigated in this work. From these conclusions, future research perspectives are identified and discussed.

CHAPTER 2

State of the Art in Wind Power Forecasting

Abstract

In this chapter, the basis of wind-power forecasting is presented. An overview of existing wind power forecasting models is presented along with the criteria commonly used to evaluate forecasting model performance. Recent developments in the areas of wind power forecast uncertainty estimation and the economic value of wind power forecasts are then reported. From these general research areas, the literature survey then concentrates on regional wind power forecasting models that have been proposed. From this review the lacks existing in the regional wind power forecasting literature are identified and the objectives of the thesis justified.

2.1 Introduction

IN this chapter an overview of existing research results in the field of wind power forecasting is presented. The aim of this chapter is not to give an exhaustive review of all existing literature on the subject. Indeed, the field of wind power forecasting calls upon many different fields such as meteorology, fluid mechanics, power systems, statistics, which all posses a vast body of knowledge. The aim of this review is rather to introduce the reader to the different aspects that constitute wind power forecasting and to illustrate the different research paths that have been explored so far. This chapter highlights the body of knowledge that the work presented in the following chapters builds upon. From this, the methods and approaches proposed in later chapters can more readily be understood and their origin can more clearly be traced.

The first section of this chapter presents the basic concepts of short-term wind power forecasting. The model for the conversion of wind power to electric power is presented. The impact of the shape of the power curve on forecasting accuracy is also explained. The input data needed, as well as the different modelling steps, are presented.

The existing criteria for model performance evaluation are then presented and their statistical meaning is explained. Indeed, quantifying the performance of a forecasting model is a central aspect of a forecaster's work. The criteria presented in this section are part of a protocol proposed in order to allow consistent inter-comparison between different models and different case studies [26]. These criteria will be used in later chapters to characterize the performance of the regional forecasting models proposed in this thesis.

The third part of the chapter presents existing "families" of wind power forecasting methods. These range from time series reference models, statistical time series approaches, physical NWP based methods, statistical NWP based methods, and combined statistical-physical methods. For each family a certain number of methods are reviewed, with an emphasis on methods that are currently in operational use by end-users. For further bibliographic results, we refer the reader to the following published state-of-the-art reviews [27, 28]. In this section, the conclusions of a forecasting "competition", which was carried out in the frame of the European Union project ANEMOS, are also presented. This competition is the first example, in the wind power forecasting field, of a true benchmarking of models. The study was co-authored by multiple modellers using a well-defined number of case studies and applying the above mentioned evaluation protocol. This benchmarking is of prime importance as it constitutes a reference for the analysis and validation of the model proposed in this work.

The fourth part of this chapter describes the advances in the literature concerning the uncertainty of wind power forecasts. Since creating a wind power forecasting model that computes perfectly accurate forecasts is conceivably impossible, providing information on the uncertainty of a forecast is useful when the forecasts are a crucial element in a decision-making process. Because wind power forecasts are less accurate than other power system forecasts such as load forecasts, uncertainty estimates are becoming increasingly sought after by end-users.

Another important aspect of any forecasting exercise is the evaluation of the usefulness or value of the forecasts. Although the accuracy of a forecasting method can be thoroughly evaluated using statistical criteria, evaluating its actual economic value is far from trivial. In the fifth part of this chapter, a review of some forecast value evaluation studies is presented.

Having presented the different fields of research in wind power forecasting, we then concentrate on regional forecasting methods. The sixth part of this chapter gives a detailed overview of existing regional wind power forecasting models. The models presented have mainly been implemented for countries where wind power has undergone important development: Denmark, Germany and more recently Spain. However, other countries, such as Ireland, which cannot rely heavily on interconnections, are highly interested in such models.

The chapter concludes with a summary of the different research areas. General conclusions are given with respect to the state of maturity of each specific area and on the current and future research trends. Concerning regional wind power forecasting, the lacks that this work aims to fulfil are clearly illustrated.

2.2 Basic concepts

A wind turbine, or wind generator converts the kinetic energy of the wind into electric energy. In this process, wind power is converted to electrical power. The amount of power a turbine can provide is directly dependent on the speed of the wind. The relation between the power output and the wind speed is modelled by the characteristic power curve of a wind turbine. Figure 2.1 depicts the typical[1] shape of such a curve. The power curve can be split into four distinctive parts. For wind speeds between 0 to 2-4 m.s^{-1} the turbine does not produce any power. The wind speed at which the turbine starts producing power is termed the *cut-in speed*. Between the cut-in speed and the *rated speed* (between 12-16 m.s^{-1}), the power curve presents a sharp slope. The shape of this part of the curve can be explained by the fact that the turbine power is a function of the cube of the wind speed [29]. The theoretical relation between wind speed and turbine power output is given by:

$$P = \frac{\rho_{air}}{2} c_p A_r v^3 \tag{2.1}$$

where P denotes the wind turbine power, ρ_{air} the air density, A_r the area swept by the rotor[2], c_p is the turbine performance coefficient, and v the wind speed at hub height. The turbine performance coefficient (the ratio of turbine power to that of wind speed) is a function of the pitch angle of the blade element to the plane of rotation and the ratio between the rotor-blade-tip speed and the wind speed. Note that according to Betz's Law, a wind turbine cannot extract more than $\frac{16}{27}$ of the power present in the wind. In practice, turbine performance ratios can reach 45%, but this does not take into account other losses such as mechanical conversion losses; modern wind turbines can extract roughly 20% to 30% of the power in the wind. The third part of the power curve corresponds to power levels close to the rated power of the turbine. The amount of power extracted from the wind is voluntarily reduced so as not to surpass the electrical generator's rated capacity. Note that in some cases the maximum power output of a turbine can be up to 20% greater than the rated power. Finally, when the wind speed reaches the *cut-off speed*, breaking mechanisms stop the turbine rotor. This is done to prevent damage to the components of the turbine due to excessive strain resulting from operation under high wind speeds.

Classical time series forecasting approaches rely on the analysis of past values of the forecast variable and of other explanatory variables to establish models. These models, given past values of explanatory variables, are very often capable of computing reasonably accurate, in the sense of useful, forecasts of the future values of a process. However, in the frame of wind power forecasting and for the forecast horizons under consideration (48

[1] Depending on the turbine model, the shape of this curve can vary slightly
[2] $A_r = \pi r^2$ where r is the rotor diameter.

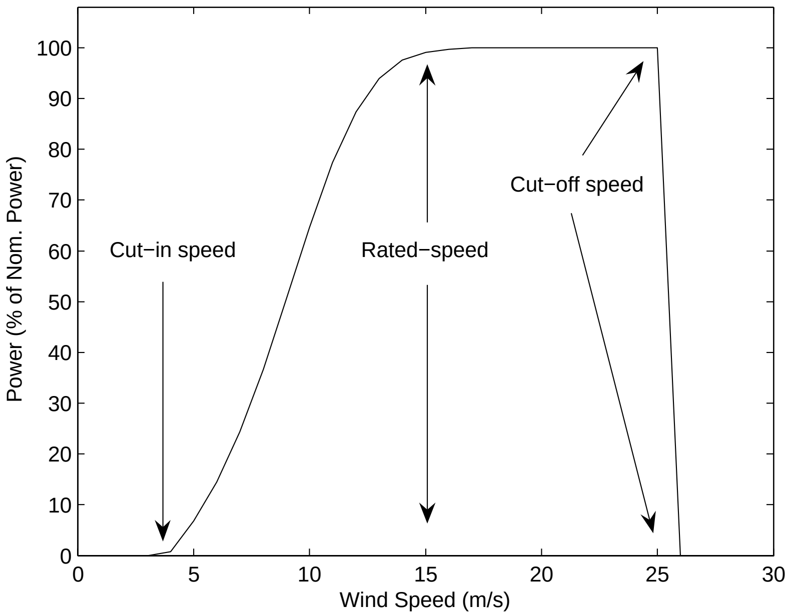


FIGURE 2.1: *Typical characteristic power-curve of a horizontal axis wind turbine.*

to 72 hours ahead) using classic time series forecasting models, such as the ARMA model proposed by Box and Jenkins [30], is not satisfactory [31] . The chaotic and non-stationary nature of the atmosphere does not allow computing useful time series forecasts of wind power for horizons longer than 4 to 6 hours [31]. To provide useful forecasts for longer horizons, the possibility that has been pursued so far has been to use NWP forecasts as supplementary explanatory variables. The basic idea of using NWP data to forecast wind power is to use the power curve to deduce the power output of a turbine or group of turbines from the meteorological variables (forecasts of wind speed, pressure, temperature, humidity, etc) provided by the weather prediction models. Although this can seem quite a straightforward procedure, several problems arise due to the nature of NWP models and the difficulty of determining a reliable power curve for a wind turbine or farm.

Numerical weather prediction models are based on the resolution of physical equations describing the state of the atmosphere. Because of computational constraints, these equations are resolved for nodes of a three dimensional grid that represents the atmosphere. This characteristic of NWP models leads to two “handicaps” with respect to wind power forecasting. The first is that a wind farm is seldom located on one of the grid nodes. Hence, the predicted meteorological variables correspond to geographical locations close to the wind farm but are not exactly those corresponding to the site. In some cases, the values provided are interpolated to the location of the wind farm. The interpolation is often performed by the NWP provider and the wind power forecaster does not always fully control the how this is performed. The second handicap of the grid used in NWP models is that they can not "see" features smaller than the grid cell size [3]. Hence, local phenomena such as hills are seldom modelled correctly, which can have a significant impact on the accuracy of the predicted meteorological variables. These approximations introduce inaccuracy in the transformation process. Given the shape of the power curve (presence of a sharp slope) these errors are amplified. This is further compounded by the fact that the sharp slope corresponds to the wind speeds most often encountered. It is generally admitted that an important part of the power forecast error is due to NWP errors [32].

Modelling the power curve of a wind generator is also far from trivial. Equation 2.1 is a simplified theoretical expression that neglects effect such as turbulence. Moreover, in the case of wind farms, the wake caused by an upwind turbine will reduce the amount of power available to a downwind turbine thus resulting in a lower than expected power output from the downwind turbine. This phenomenon is typically referred to as shadowing effect. Although wind turbine sitting within a farm is planned in order to minimize this effect, it can profoundly affect the power output of a wind farm depending on the direction of the wind.

Following the above, wind power forecasting can be broken down into two stages. The first, termed *downscaling* in the wind power forecasting community, deals with the scaling of the NWP wind speed to the wind speed at the generator's hub height or at the location of the wind farm. The second stage, called *power curve modelling*, corresponds to the transformation of the downscaled wind speed into power, for a wind turbine or a wind farm. In the following sections, models, which follow this basic two-stage concept, will be presented. Some models, such as physical models, explicitly perform both steps. Others, such as statistical models, often combine both steps in a single computation.

2.3 Model Performance Evaluation

In forecasting, determining which forecasting method is the best is one of the forecaster's prime concerns. This however, can be a misleading problem. As stated in [33], a simple method, which is only marginally less accurate than a more complicated or costly one, will most often be preferred in an operational setting. In most cases, model performance is determined by examining model accuracy. To determine model accuracy, though, is also a complex issue, given that several model accuracy measures exist. In this section, we will describe the error measures most often employed in the wind power forecasting community. These measures have been included in a protocol for short-term wind power model performance evaluation proposed in [34]. This protocol was proposed to help end-users as well as researchers compare model performance in the frame of the important development of the short-term wind power forecasting field.

A model's prediction error is classically defined as the difference between the measured and the predicted value. Because short-term wind power forecasts are usually provided for several time-steps in the future (typically the next 48-72 hours with hourly time-steps), a

[3]Model grid sizes vary between different NWP model implementations. In the frame of operational wind power forecasting the resolutions usually vary between $0.5^\circ \times 0.5^\circ$ and $0.15^\circ \times 0.15^\circ$.

horizon dependent model error $e(t+k|t)$ should be considered:

$$e(t+k|t) = P(t+k) - \hat{P}(t+k|t) \tag{2.2}$$

where $P(t+k)$ is the measured power at time $t+k$, $\hat{P}(t+k|t)$ is the power forecast for time $t+k$ computed at time t. In order to compare model performance for different wind farms, it is convenient to use a normalized prediction error. Some results in the literature normalize the error based on the observed power generation. However, as stated in [35], this leads to exaggerating the error for low wind power generation. For this reason, the error is most often normalized by the installed capacity:

$$\epsilon(t+k|t) = \frac{1}{P_{inst}}\Big(P(t+k) - \hat{P}(t+k|t)\Big) \tag{2.3}$$

where P_{inst} is the nominal power of the considered wind farm.

The prediction error, when seen as a time series can be decomposed into a systematic and a random component:

$$e = \mu_e + \varepsilon_e \tag{2.4}$$

where μ_e is the constant or systematic error and ε_e is a random variable. The systematic error is also called the model bias and can be estimated for each horizon using:

$$\mathrm{BIAS}(k) = \hat{\mu}_e(k) = \overline{e(k)} = \frac{1}{N}\sum_{t=1}^{N} e(t+k|t) \tag{2.5}$$

where N is the number of model runs over which the model is evaluated. Note that model bias should be as close to zero as possible. If this is the case, the forecasts can be said to be unbiased.

The two most commonly used evaluation criteria are the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE). The MAE can be defined as:

$$\mathrm{MAE}(k) = \frac{1}{N}\sum_{t=1}^{N} |e(t+k|t)| \tag{2.6}$$

where $|.|$ denotes the absolute value. The RMSE can be derived from the Mean Square Error (MSE) defined as:

$$\mathrm{MSE}(k) = \frac{\sum_{t=1}^{N}(e(t+k|t))^2}{N} \tag{2.7}$$

Hence, the RMSE is defined as:

$$\mathrm{RMSE}(k) = \sqrt{\mathrm{MSE}(k)} \tag{2.8}$$

$$\mathrm{RMSE}(k) = \sqrt{\frac{\sum_{t=1}^{N}(e(t+k|t))^2}{N}} \tag{2.9}$$

It can be noted that both the systematic and random errors influence the values of the MAE and RMSE. To study the influence of the random error, the Standard Deviation of Errors (SDE) can be computed:

$$\mathrm{SDE}(k) = \hat{\sigma}_e(k) = \left(\frac{\sum_{t=1}^{N}(e(t+k|t) - \overline{e(k)})^2}{N}\right)^{\frac{1}{2}} \tag{2.10}$$

This criterion measures the standard deviation of the error distribution for a given horizon.

From a statistical standpoint, the BIAS and the MAE are linked to the first moment of the prediction error distribution; they are related to the power output and hence are usually the most easily interpreted from an end-users perspective. The RMSE and SDE are associated to the variance or second moment of the error distribution. Although not as easily interpreted from the end-user perspective, the fact that these measures strongly penalize large errors makes them significant as large prediction errors are the most costly to deal with. Note also that all the above-mentioned error measures can be computed using the prediction error $e(t+k|t)$ or the normalized prediction error $\epsilon(t+k|t)$. In the latter case, the notation of the error measures becomes: NBIAS, NMAE, NRMSE and NSDE; where the N stands for normalized. Using the normalized error measures is useful to compare performance among wind farms of different nominal capacities.

As suggested in [36], the skewness and kurtosis permit a distribution-oriented analysis of the forecast error. The skewness can be estimated using Fisher's formula:

$$\hat{\nu_e}(k) = \frac{N}{(N-1)(N-2)} \sum_{t=1}^{N} \left(\frac{e(t+k|t) - \hat{\mu}_e(k)}{\hat{\sigma}_e(k)}\right)^3 \tag{2.11}$$

The skewness is linked to the third moment of a distribution; it indicates the degree of symmetry of the error distribution. If the skewness is null then the distribution is symmetrical, if the skewness is negative then the distribution is left-skewed (the left tail of the distribution is the "longest") and, if the skewness is positive, the distribution is right-skewed (the right tail of the distribution is the "longest").

The excess kurtosis can be defined as:

$$\hat{\kappa_e}(k) = \frac{N(N+1)}{(N-1)(N-2)(N-3} \sum_{t=1}^{N} \left(\frac{e(t+k|t) - \hat{\mu}_e(k)}{\hat{\sigma}_e(k)}\right)^4 - \frac{3(N-1)^2}{(N-2)(N-3)} \tag{2.12}$$

The excess kurtosis is linked to the fourth moment of the distribution. It provides information on the shape of the error distribution with respect to a Gaussian distribution. If the kurtosis is positive, the distribution presents a sharper peak around the mode and longer tails than the Gaussian distribution. Conversely, if the kurtosis is negative the distribution presents a flatter peak around the mode and shorter tails.

Another performance measure suggested in [34] is the coefficient of determination, denoted R^2. Several definitions for this criterion exist, the most common being the square correlation coefficient between the measured and predicted wind power. However, the authors argue that with this definition, $R^2 = 1$ can be obtained although the forecasts might be biased. The authors therefore propose the following definition be used:

$$R^2(k) = \frac{\mathrm{MSE}_0(k) - \mathrm{MSE}(k)}{\mathrm{MSE}_0(k)} \tag{2.13}$$

where $\mathrm{MSE}_0(k)$ is the Mean Squared Error obtained when using the climatological mean model (see Equation 2.17)

$$\mathrm{MSE}_0(k) = \frac{\sum_{t=1}^{N}(P(t) - \overline{P(t)})^2}{N} \tag{2.14}$$

This coefficient gives an estimate of the model's ability to reflect the variance of the data. The coefficient's values lie between 0 and 1, where 1 denotes perfect prediction (in that case the MSE is null). Note that this criterion does not directly measure model accuracy but rather compares an advanced model's performance to that of the climatological mean model. For this reason the authors argue against using it as the main model performance evaluation criterion. However, this criterion can be useful to compare different models, or the performance of one model for different data sets. The final recommendation of the authors concerning this criterion is that given the different existing definitions for R^2, the definition being used should be clearly stipulated when reporting model performance results.

In the frame of model comparison, especially when benchmarking advanced models with respect to reference ones, the definition of the coefficient of determination given in Equation 2.13 can be generalized. This generalization encompasses both the reference model and the evaluation criterion. This generalization basically measures the improvement, in terms of an evaluation criterion, of one model with respect to another model:

$$Imp_{ref,\mathrm{EC}}(k) = \frac{\mathrm{EC}_{ref}(k) - \mathrm{EC}(k)}{\mathrm{EC}_{ref}(k)} \tag{2.15}$$

where EC denotes the evaluation criterion and ref the reference model considered. The evaluation criterion can be any of the above-mentioned criteria: MAE, RMSE, SDE, or their normalized versions.

In the following chapters the evaluation criteria presented in this section will be used to assess the accuracy of the regional wind power forecasting models.

2.4 Overview of state-of-the-art wind power forecasting models

In this section, an overview of existing wind power prediction methods is presented. Emphasis is given to the classification of the methods and general characteristics are presented for the different sub-families.

In general, models can be classified as either involving Numerical Weather Predictions (NWP) as input or not. With regard to methods that incorporate NWP data, two mainstream approaches exist: the physical and the statistical approach. In some models, a combination of both is used, as indeed both approaches can be needed to provide useful forecasts.

The physical methods use NWP and physical considerations to reach the best possible estimate of the wind speed at the location of the wind farm. A power curve is then used to convert the wind forecast into the power forecast. Most models in this class can use Model Output Statistics (MOS), which is a statistical correction, to reduce the remaining systematic error. In case a MOS is used, the methods also rely on past production measurements.

Statistical models try to find the relationship between a number of explanatory variables including NWP forecasts, and measurements of past power production or meteorological variables. The statistical methods can be seen as regression models that try to estimate the parameters of a function linking future wind power to available explanatory data.

As mentioned in section 2.2, time series approaches, in the sense of using only data from the past to forecast the future, have consistently been shown to under-perform methods that incorporate NWP data for horizons between 6 to 48 hours. Although these horizons are the main focus of this work, a short review of existing results based on the time series approaches will be given, as these methods often perform quite well for the short very term.

2.4.1 Time series Models

Reference Models

In forecasting, implementing an advanced, often costly, model is only worthwhile if it outperforms simpler models. With this aim, forecasters have defined a number of simple models, which can be used as benchmark models.

The first model in this class is the Persistence model or naïve predictor. In this model, the forecast for all times ahead is the current value. The Persistence model is given by:

$$\hat{P}(t+k|t) = P(t) + e(t+k) \tag{2.16}$$

where $\hat{P}(t+k|t)$ is the forecast computed at time t for horizon $t+k$, $P(t)$ is the wind power measured at time t and $e(t+k)$ is the residual. Hence, the error for zero time steps ahead is zero. In wind power forecasting, this model usually performs well because the time scales of atmospheric changes are in the order of several hours. For short forecasting horizons (0-6 hours ahead), this model is considered in the literature as the benchmark all other predictors have to beat.

The Persistence model is a particular case of the more general moving average model, which can be expressed as:

$$\hat{P}(t+k|t) = \frac{1}{n}\sum_{i=0}^{n-1} P(t-i) \tag{2.17}$$

where n is the order of the model, and specifies over how many past values the average is computed, hence for Persistence the model order is $n = 1$. Further as n tends to infinity, this model becomes a global average or climatological mean:

$$\hat{P}(t+k|t) = \overline{P(t)} \tag{2.18}$$

where $\overline{P(t)}$ is the average of all available power measures at time t.

For the specific case of short-term wind power forecasting, a new reference model has been proposed [37]. This model is basically a weighted sum between Persistence and the mean power production:

$$\hat{P}(t+k|t) = a_k P(t) + (1-a_k)\overline{P(t)} \tag{2.19}$$

where a_k is a horizon dependent weighing factor. The idea behind this model is to couple the short-term "strength" of persistence with the long-term one of the climatological mean. For a sufficiently large horizon k, the production at time t and at time $t+k$ are uncorrelated [37]. Because of this, using persistence for large horizons is equivalent to predicting a random wind power value. Therefore, predicting the mean production results in better forecasting accuracy in terms of the root mean square error for large horizons. The authors of this method suggest using the correlation coefficient between $P(t+k)$ and $P(t)$ as the value for a_k. They investigated the possibility of defining a_k as a function of the correlation for $k = 1$:

$$f(k) = \phi^k \tag{2.20}$$

where ϕ is the estimated correlation for horizon 1. However, they found large deviations between $f(k)$ and the correlation, especially for summer months where correlation is periodic due to diurnal variations. For a wind farm in Denmark, this new reference model outperforms persistence for horizons superior to 3 hours and outperforms the mean for horizons up to 24-36 hours ahead. This reference model is adequate for all forecast lengths and, like Persistence, requires only measured time series as input. However, it does no seem to have been widely adopted in the literature. This disaffection might be linked to the significant complexity difference between implementing persistence and implementing this model.

For the special case of regional forecasting, the total regional power is seldom available on-line. Therefore, a model such as Persistence cannot be used because it is based on the use of the current measure of power. Furthermore, comparing regional forecasting models to an "artificially" computed Persistence in an off-line evaluation setting would lead to an underestimation of the regional model's performance. To overcome this problem the *on-*

line persistence has been proposed in [38]. This model is defined as:

$$\hat{P}^{reg}(t+k|t) = \frac{P_{inst}^{reg}}{\sum_{i=1}^{M_r} P_{inst,i}^{wf}} \sum_{i=1}^{M_r} P_i^{wf}(t) \tag{2.21}$$

where P_{inst}^{reg} is the total installed capacity in the region, $P_{inst,i}^{wf}$ is the installed capacity of the M_r wind farms for which on-line measures are available, and $P_i^{wf}(t)$ is the power measured at those wind farms at time t.

Time series models

The reference models described above are in general difficult to beat for very short-term horizons. In some applications such as control of medium-sized island systems, persistence-like models (in this case a moving average) were found to be optimal [39].

Statistical models such as ARMA, ARX and Box-Jenkins methods have been widely used for time series prediction, either for finance, control, or other applications [40]. They have been applied to the wind power forecasting problem [31, 41] with some success especially for very short-term forecasting (0-6 hours ahead).

In [42] autoregressive models were used to predict wind power for the next 6, 12 and 24 hours. The authors found that these models consistently outperformed Persistence, with up to 15% improvement with respect to mean error. They also showed that concerning the training period, one year is enough to correctly estimate the model parameters.

Similar results were found in [43] using a linear autoregressive model and an adaptive fuzzy logic based model. Although the results also beat Persistence by a significant margin (20% improvement based on root mean square error), they showed that for horizons superior to 8 hours, the 95% confidence interval of the spot predictions contained most of the likely values. Hence, the values predicted for “medium” and “long” term horizons cannot be trusted to a great extent.

An alternative to classic statistical models is neural-network based models. These models are trained over long collections of production data using specific learning algorithms like the back propagation algorithm [44, 45]. Neural networks are interesting because of their flexibility. Different network architectures and learning schemes can easily be tried [25, 46]. In [47, 48] the author compared several models of this family for the problem of wind speed forecasting, giving a brief description of the manner in which these models were designed.

2.4.2 Physical Models

Physical models differ from statistical ones in that they try to use physical laws to downscale the NWP forecasts to the local wind conditions over a wind farm and then use power curves to transform the predicted wind into predicted power. To perform the downscaling, physical approaches use meso-scale or micro-scale NWP models. It must be noted that these downscaling models require detailed physical descriptions of the wind farms and their surroundings: wind farm layout, roughness, orography, etc. Acquiring these data is one of the main drawbacks of physical models [27]. However, contrary to statistical models, physical models do not necessarily need historical data to be implemented.

Physical downscaling models are already fully tested state-of-the-art models and can run for several hours (7-8 hours for the Local Circulation Assessment and Prediction System model [49]). Some models use commercial codes like STAR-CD to solve the local effects on the wind field [50]. These Computational Fluid Dynamics codes are mainly used for complex terrain situations, either to carry out wind power forecasting or wind resource assessment. They can also be used to study the turbulent energy over cliffs [51].

The Prediktor model, developed at Risø National Laboratory is a physical model that uses the WasP (Wind Atlas Analysis and Application Program [52]) methodology to convert the forecast wind from NWPs (HIRLAM in this case) into the wind seen by the turbine. The turbine power curve is used to obtain the wind power [53, 54]. This model is now part of the Zephyr tool with the WPPT model developed by IMM [55]. The aim of Zephyr is to merge these two models to obtain a synergy between the physical and the statistical approaches.

The University of Oldenburg [56] has developed a very similar approach but with the Deutschlandmodell instead of HIRLAM for the numerical weather predictions. The model is named Previento [57].

LocalPred is a tool developed by CENER, the Spanish National Renewable Energies Center, in collaboration with the Spanish Research Center for Energy, Environment and Technology (CIEMAT). It involves adaptive optimization of the NWP input, time series modelling, meso-scale modelling with MM5, and power curve modelling [58].

A company in the United States, TrueWind Inc. [59] has developed the eWind model that runs a numerical weather model as a meso-scale model (ForeWind) using boundary conditions derived from a regional weather model. More processes that are physical are captured and the prediction can be tailored better to the local site. This model is commercially available.

2.4.3 Statistical Models

The Institute for Informatics and Mathematical Modelling (IMM) of the Technical University of Denmark developed the Wind Power Prediction Tool (WPPT). It has been in operation in the control rooms at ELSAM and ELTRA (system operators for the Jutland/Funen area in Denmark) since 1994 [60]. WPPT is a conditional parametric model that has regularly been improved during the last decade. HIRLAM forecasts were integrated to the model [61], which allowed the range of useful forecasts to be extended to 39 hours ahead.

More recently, a fuzzy-logic based approach has been developed. These models were first designed for control purposes [62]. They appeared to be suitable for solving a large range of problems and were easily applicable to forecasting. The resulting models have been used for both load forecasting and wind power forecasting [63, 64]. These fuzzy models use if-then rules and are trained, like neural networks, over long series of known production data. A simulated annealing algorithm controls the learning process and cross-validation is applied to terminate learning [64]. This is the type of model used in the Armines Wind Power Prediction System (AWPPS) developed at the Center for Energy and Processes (CEP).

The strong wind energy growth in Spain led Red Eléctrica de España (the Spanish TSO) to have the Sipreólico tool developed by the University Carlos III of Madrid [65]. The tool is based on Spanish HIRLAM NWP forecasts, taking into account hourly on-line power measurements provided by a Supervisory Control And Data Acquisition System (SCADA). These inputs are then used in adaptive non-parametric statistical models, together with different power curve models. There are 9 different models, depending on the availability of data. The results of these models are then used in a forecast combination scheme to compute the final forecast.

A model named GH Forecaster has been developed by British company Garrad Hassan and Partners [66]. This converts NWP forecasts and on-site measures using a classic two-stage approach. The NWP forecasts are downscaled to the level of the wind farm using adaptive multi-parameter statistical regression techniques. The refined wind speed forecasts are then converted to power forecasts using direction dependent conversion matrices. These matrices can be derived from theoretical data or constructed statistically from past measurements.

2.4.4 Model benchmarking

In the frame of the EU project ANEMOS, several of the above-mentioned models (Prediktor, WPPT, Previento, Sipreolico, LocalPred, AWPPS to name a few) were benchmarked on several different wind farms across Europe. The results of this benchmarking exercise were published in [32]. Since this is probably the first benchmarking of the most advanced European models and since the models compared were both physical and statistical, the main conclusions of the study are presented here. The comparison of models was made possible by a standardization procedure of the available data and the definition of an evaluation protocol [34].

The analysis of different NWP models available for the benchmark showed that they could provide useful wind predictions for wind power forecasting. Further, spatial resolution is of major importance especially in complex terrains; low spatial resolution models lead to higher errors in wind predictions. Finally, post treatment of NWP using statistical tools such as Kalman filters was demonstrated to improve NWP accuracy [67].

The main conclusion of this benchmark is that the performance of prediction models is highly dependent on the complexity of the terrain where the wind farm is located. In flat terrains, the mean absolute error of the models for the 24 hours horizon was below 10% of the nominal power of the wind farms, whereas in complex terrain this error surpassed 35%. One of the test cases was the Tunø Knob offshore wind farm. For this farm the performance of the models was similar to that found for flat terrain.

Finally, the performance of the models was not "consistent" for all cases and all horizons; no model performed best for all horizons or all farms. A combination of the forecasts could therefore be studied to achieve the best overall performance for all horizons and all wind farms. However, the statistical models showed the most consistency and have a slight general advantage over physical models especially in the more complex terrain types.

The results of this benchmarking will be used in later chapters in order to validate the performance of the RPC forecasting model proposed in this thesis.

2.5 Uncertainty of Wind Power Forecasts

The prediction methods described so far provide point predictions for a number of hours ahead. Although the performance of these methods can be evaluated using the criteria presented in section 2.3, these criteria only give end-users an idea of the average performance of the models. In an operational setting, where decisions are to be based on the forecasts, knowing that a model is usually good, is not enough. Because of this, end-users have expressed the need for tools that not only provide the expected production value, but also provide an estimate of the accuracy of a given forecast run. In other words, the users want to know how much confidence they can have on the forecast.

The first response to this need proposed in the literature was to provide prediction intervals. Prediction intervals give an estimate of where the production will lie with a certain degree of confidence. An example of such a forecast is given in Figure 2.2, where the prediction intervals are represented by differently coloured areas. The solid line with rhombi represents the power measurements corresponding to the hours for which the forecast was provided. The solid line represents the points forecasts associated with the prediction intervals.

A parametric model proposed in [68], models the conditional distributions of *wind speed* forecasts error (given the predicted wind speed) as Gaussian distributions. The derivative of the wind farm power curve is then used to transform the Gaussian distributions of wind speed error into non-Gaussian distributions of the power forecast error. The prediction intervals can then be derived from the estimated power error distributions.

Another parametric model is proposed in [69] where β-distributions, bounded between 0 and 1 in the same way as normalized forecast errors, are used to model the forecast error distributions. The mean of the predictive distributions is set as the point forecast. The variance parameters are determined by studying the historical performance of the considered forecasting method.

In [36], the author proposes a non-parametric algorithm that estimates the intervals from a set of past error values. The intervals are computed using conditional distributions of power forecast error (given predicted wind speed). The error and wind speed data over which the distributions are computed are determined using a fuzzy inference system. Once this data has been selected, a resampling method is used to determine the desired distributions and from them the interval forecasts. The advantage of this method over previously proposed parametric methods is that it does not rely on a hypothesis on the type of error

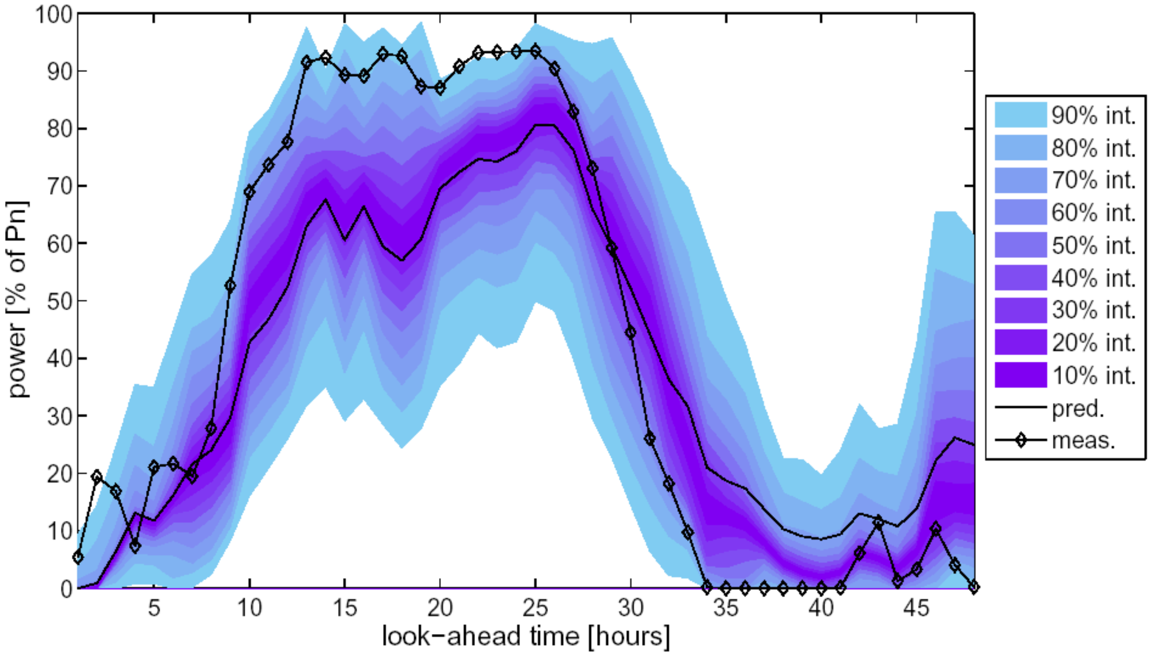


FIGURE 2.2: *Example of wind power point predictions associated with a set of interval forecasts. This example is taken from [36].*

distribution.

Another class of non-parametric methods has been proposed for interval forecasting: quantile regression. As its name implies, the method estimates predefined quantiles from a set of input data. In [70], the proposed method uses the point forecasts provided by another prediction model as well as explanatory data used by the point forecasts model. This method can therefore be added to an existing installation to provide prediction intervals for the point forecasts. In [71], the author proposes a different method to compute prediction quantiles directly from the NWP forecasts. One drawback of these methods is that a specific model has to be tuned for each quantile, hence for each desired prediction interval.

In complement to interval prediction methods, which are based on "traditional" NWP data that provide a single value for a given horizon, ensemble prediction methods are designed to take advantage of ensemble meteorological forecasts. Ensemble NWP provide several likely values for a given horizon. For example, the different values are obtained by running differently calibrated NWP models, or by slightly modifying the inputs to the NWP models, in order to obtain a set of different forecasts. The rationale behind ensemble forecasts is that NWP results are highly dependent on the initialization and parameterization of the NWP model. Thus, by running the NWP model with slightly differing parameters or inputs, different results (or members) are obtained which, theoretically, represent equiprobable outcomes. Hence, the spread over which the ensemble members lie can theoretically provide an indication of the uncertainty of the forecast.

One way to obtain ensembles when none are available is to use so called poorman's ensembles. These ensembles are derived from "classic" NWP forecasts; the "ensembles" are created by considering consecutive overlapping model runs. In [72], such ensembles are used to compute a Meteorological Risk Index, which is intended to give end-users a clear indication on the trustworthiness of a given wind power forecast. In [36], the author further extended the MRI method to use true NWP ensembles. The results show good consistency between the index value and the level of forecast error.

Another aim, which is currently pursued, is to derive probabilistic forecasts of wind power from ensemble NWP data. The purpose of these methods is to forecast the future probability distribution of wind power. An example of the research, which has been conducted in this area, can be found in [73], where forecast quantiles are derived from NWP ensembles computed by the National Centres for Environmental Prediction (NCEP) and the European Centre for Medium-Range Weather Forecasts (ECMWF). In this study, the Zephyr/Prediktor wind power forecasting model, as well as the power curve model approach described [74], are used. This study showed that using a biased power curve, which is normally used for point predictions, leads to inaccurate quantiles. The unbiased curve should span all possible power values so that the quantile forecasts are probabilistically correct.

2.6 Value of Wind Power Forecasts

Although the importance of wind power forecasting is undisputed as a means to ease the integration of large amounts of wind power into power systems, relatively few studies have been conducted to asses the economic value of wind power forecasting to utilities. One of the main reasons for this is the uniqueness of power systems. Each system has a different generation mix; hence, each system's sensitivity to the variable and uncertain nature of wind power will be different. A system possessing many fast response dispatchable units will be less sensitive to wind power variations than a system that relies more heavily on slower base-load units. Another issue to be taken into account is the wind regime and the distribution of the wind parks. A system where the wind parks are spread over a wide area will face a more stable aggregated wind power production, than, for example, an island system where the production of all wind parks is highly correlated. Another factor which has been increasingly put forward as hampering wind development is system interconnection. A well-interconnected system such as western Denmark can "easily" manage large quantities of wind power and important unforeseen variations [75] by importing energy from neighbouring systems. Isolated or lightly interconnected systems, such as the Irish one, can less easily handle severe forecast errors, thus limiting the amount of wind power capacity that can be connected to the grid.

In [76] a power system model was used to investigate the impact of wind power forecast accuracy on the cost of system operation. The Elfin probabilistic cost model, which, given forecast load and data on the production means, calculates the expected generation from each unit, was used to compute the operation cost for three different utilities. For each utility, the authors varied the accuracy of the simulated wind power forecasts that were treated by Elfin. The results of this study show, not surprisingly, that increased accuracy leads to lower operating costs. However, the penalty associated to over or under-estimating future wind power is not symmetric; the cost associated to a given positive error will not be equal to the cost associated to the opposite negative error. Furthermore, the authors found that this dissymmetry can vary from one utility to another. For one utility over-estimating was less costly while for another utility the opposite was true.

Another difficulty in evaluating the usefulness or the benefit derived from wind power forecasts is the difficulty of knowing exactly how the forecasts are used to manage the power system. In [73] the authors surveyed the use of a forecasting system at Elsam (a Danish utility) and found that the use varied among employees. Some referred to the forecasts very often whereas others only used the forecasts to assess certain situations. Modelling control-room use of wind power forecasts is certainly one of the most difficult aspects of forecast value evaluation.

Given the difficulty of directly estimating the reduction of power system operating cost linked to wind power forecast error, many studies have taken a slightly different approach by studying the value of wind power forecasts as trading tools in the electricity market. The relatively recent move to liberalize the electricity markets in many countries has come at the same time as the widespread development of wind power, especially in Europe. The rationale behind these studies is that prediction error costs can "easily" be linked to the imbalance penalties to which wind farm operators are submitted when they trade electricity in the market. In a sense, the imbalance penalties can be considered as reflecting, in part, the added regulation cost faced by a power system dealing with the variable nature of wind power.

One such study was conducted for the Nordpool market [77]. The authors simulated an important amount of wind power production in Denmark and examined its impact on the Nordpool. They showed that using a power forecasting model (in this case the one developed at Risø National Laboratory) with the standard Nordpool rules (24 hour bids starting at midnight, bids submitted at noon) or modifying the rules so that trading can be continuous and using linear models (3 hour ahead forecasts) resulted in imbalance penalties corresponding to roughly 12% of the average Pool price.

In [78] the case of the Spanish electricity market was examined. The authors propose a probabilistic methodology to estimate the cost of wind prediction errors. They modelled the forecast errors using β-distributions following the work described in [69, 79]. Based on this model of the errors (i.e. the α and β parameters derived from the mean and standard deviation of the errors) and on the regulation prices, the authors calculate the probable penalties that would be paid by the wind power producers over a certain period. In this study, conducted for a single wind farm, fifteen wind farms and the total wind power production of Spain, the authors used mean and standard deviation values obtained for northern Germany by the Previento model [24, 79]. The study results show that the annual prediction errors would cost roughly 10% of the farm's income from the electricity market. Further, the authors showed that the cost of error varied over the year, with higher prices in the summer months. In addition, they found that bidding with aggregated production from a regional or national perspective led to lower error costs.

In another study described in [36, 80, 81] the case of the Dutch APX electricity market was examined using the AWPPS model developed by Kariniotakis [25] to provide wind power forecasts. The authors found that by using the model alone the earning represented 87% of the earning which could have been obtained by using perfectly accurate forecasts, that is forecasts with zero error. The authors further proposed to use the uncertainty assessment methods described in [82] to develop specific bidding strategies that take into account the asymmetric nature of the imbalance prices. Using these methods, they obtained earnings of up to 96% of those possible using perfect forecasts. Hence, by using the added information on the uncertainty of wind power forecasts, it is possible to enhance the economic value of wind power forecasts without improving their absolute accuracy.

2.7 Regional Prediction Models

Regional and national wind power forecasting models are mostly based on extrapolation also known as upscaling. Extrapolation means that to predict the regional or even national power from on-line measurements and NWPs, data from only some wind farms spread in the region or country are used. Indeed, obtaining power measurements and NWP data from all the wind farms in a region or country is very expensive and can be almost impossible if hundreds of wind farms are considered. This problem has mainly been pointed out in Germany in the last few years [56] but every country that wishes to increase its wind energy penetration rate will be concerned in the near future.

In Previento (a tool developed at Oldenburg University) [57] an extrapolation algorithm is included. This model uses the physical model described earlier to compute single wind farm forecasts for representative wind farms in the region. The extrapolation of the single wind farm forecasts to the regional forecast is based on the correlation between the representative wind farm productions and the total production computed on past measurements. The root mean square error for the regional predictions, normalized to the installed power capacity, is between 4% and 7% depending on the forecast time.

In [24, 83] the authors wrote that predicting the regional wind power produces less error than predicting the wind power at single sites because of spatial smoothing effects. They estimate that the magnitude of the error reduction strongly depends on the size of the region and the number of production sites it contains. Their conclusion is that only a few representative sites are necessary to predict the wind power production for a whole region with a reduced error.

From a recent conversation with this model's developers, U. Focken and L. von Bremen, it appears that a novel upscaling algorithm based on neural network is under development. Further, representative wind farm selection will be based on physical considerations i.e. representative wind farms will be selected to present average roughness lengths and average power curves similar to those of other wind farms in their neighbourhood. Finally, it appears that although good quality on-line regional production measurements are available (measurements of 80% of the regional power available once daily) they will not be used for on-line adaptation of the regional prediction model.

The upscaling approach developed at IMM is different [84]. The predictions for a region are calculated using a two-branch model. In the first branch, single wind farm predictions are computed for a number of reference wind farms. These individual predictions are then combined and upscaled by an upscaling model. The second branch of the model uses area prediction models to compute forecasts for sub-areas of the region. The area model uses area NWPs (typically averages of the NWPs available for the wind farms in the sub area) as well as the sum of measured production from the reference wind farms in the area. The forecasts produced by both branches are then combined using a weighted, horizon dependent, average which uses the root mean square error as the weighing criterion. The models in both branches are conditional parametric models similar to the last version of the WPPT model.

This model has on-line adaptation capabilities because all the models it uses have this capability. The single wind farm models can obviously be adapted since reference farms almost always provide on-line measurements of their production. The upscaling models can be adapted with a delay if a measure of the regional production is available. This, however, is rarely the case in practice. For example, regional measures are only available with a few weeks delay in the case of the Danish utility ELSAM. Considering this, the model can be adapted, but the way in which this might be done is not clear. With respect to the sub-areas used in this model it appears, from a conversation with the model developers, that their definition was mainly the result of end-user requirements. In a similar manner, the reference farms where selected primarily on technical considerations such as reliable SCADA systems and nominal power.

ISET (Institut für Solare Energieversorgungstechnik) has operationally worked with short-term forecasting since 2000, using NWP provided by the German National Meteorological Service (DWD) model and neural networks. Their first application was for E.On, who initially lacked an overview of the current wind power production and therefore wanted a good tool for nowcasting [85]. The nowcasting model was later adapted to provide longer-term forecasts.

Rohrig [86] reported on the latest developments of ISET's Wind Power Management System WPMS. In the case of E.On Netz, the 5.7 GW installed capacity are upscaled from 50 representative wind farms totalling 1850 MW. The model performs in two steps: individual wind farm forecasts are first computed and then aggregated and upscaled by a transformation model. The selection of the representative wind farm sites is based on the modeller's experience. From conversations with the model developers, it appears that the ANNs used in this model do not integrate on-line adaptation capabilities.

The extrapolation model developed at the Centre for Energy and Processes is based on fuzzy-neural networks. The models described in [38] were evaluated for the case of Ireland.

Another regional prediction model is RegioPred [58]. This model is based on the single wind farm prediction model LocalPred, which uses high-resolution Computational Fluid Dynamics (CFD) models and statistical power curve modelling to compute the power forecasts. The reference wind farms used in this model are selected using cluster analysis.

In [35, 87] an ensemble prediction model is used to forecast the aggregated production of Ireland, Germany and Western Denmark. This model is based on the Multi-Scheme Ensemble Prediction System (MS-EPS), proposed by Möhrlen [14]. The MS-EPS is a 75 member, limited-area model that is commercially operated by WEPROG. This system can provide ensemble power forecasts for the next 48 hours. The way in which the regional forecasts are computed is not clear. For single wind farms, the model uses direction dependent power curves, which are calibrated for each ensemble member.

In [88] the existence of a commercially available regional forecasting model: Simulation Model for the Operational Forecast of the Wind Energy Production in Europe (a.k.a.: SOWIE), is reported. This model, developed by German company Eurowind GmbH, is a physical model that relies on HIRLAM NWP forecasts and a database of wind turbine power-curves to compute individual forecasts for all wind turbines in Germany. From these forecasts, the total wind power forecast is derived for horizons up to 120 hours ahead.

More recently, a statistical model very similar to that implemented in WPPT has been proposed in [89, 90] for northern Japan. In this approach, the authors propose a two-step conversion of NWP data into power forecasts. They first downscale the 20 kilometre resolution NWP forecasts provided by the Japanese weather service using a direction dependent scaling factor. This statistical downscaling approach is tuned using the 20 km resolution NWP data and one year of high resolution (1 kilometre) data computed using the meso-scale model RAMS. The downscaled forecasts are then converted into power using a model named Auto-Regressive with Exogenous input and Multi-timescale parameter model (ARXM). This model is very similar to that described in [91] with an added parameter to take into account non-measured wind farm outages. To derive regional forecasts from the reference wind farm forecasts the authors propose using an upscaling factor based on the ratio between non-reference wind farm load factors and reference farm load factors as described in [92].

In [93] the authors propose a different approach where a "physical" approach is combined with a statistical one in order to forecast the total German wind power production. In this approach, a "physical" conversion of forecast wind speed to power is computed for 306 subs-regions. The conversion is performed by multiplying the sub-region's installed capacity by the square of the forecast wind speed. This provides a meteorologically dependent power production estimation. This estimation is computed on historical data spanning several months and then principle component analysis (PCA) is performed by considering the values obtained for the 306 sub-regions as variables. The six first principle components are then used in a multi-linear regression model to determine the relation between a meteorologically dependent situation and the total production. To derive the forecasts, the NWP runs are used to derive the 306 sub-regional power values, which are then projected on the first six dimensions determined by the PCA on the historical data. The projected values are then fed into the multi-linear model to obtain the final total production forecast.

2.8 Conclusion

Wind power forecasting is a complex problem involving different fields: meteorology, physics, statistics and power system engineering. The developments in this field have followed the needs expressed by the industry. Prediction models have evolved from short term forecasting for wind turbine control, to wind farm forecasting and now to regional forecasting. The more recent developments in forecasting have mainly dealt with forecasting the power output of wind farms or the aggregated power output of wind farms in a given region. The dominant type of wind power forecast is currently based on the refinement and conversion of NWP forecasts and on-line measurements into power forecasts. Several approaches, both physical and statistical, have been implemented and are now commercially available.

Most operational methods today present similar performance. They all perform better when benchmarked on flat rather than complex terrains. Most wind power specialists agree that the principal source of power forecast error stems from inaccurate NWP forecasts. Meteorologists are actively pursuing research to improve the accuracy of NWP models, and a promising short-term solution to improve the usefulness of wind power forecasts appears to be the use of meteorological ensembles.

Regional forecasting has become a necessity for TSOs or utilities wishing or facing an increase in wind power penetration in their system. For large wind farm operators trading in the new liberalized electricity markets where aggregated bids are possible, having a forecast of the total production is also important. Further, it has been shown that regional forecasts benefit from a smoothing effect and are thus more accurate than single farm forecasts, relatively to the installed capacities under consideration.

From the literature available on the subject of regional forecasting some points appear. So far, studies have mainly concentrated on cases where the amount of installed capacity was very large: Germany and Denmark. Yet, for smaller and less interconnected systems such as Ireland and Crete, interest in regional prediction is apparent because of high or increasing wind power penetration. Few studies have looked into the particularities of regional forecasting for this type of power system. Most existing regional forecasting systems were developed to answer specific end-user requirements. Little has been published regarding general design methodologies to implement regional forecasting models. The aim of this thesis is to study regional forecasting from a general point of view and to establish guidelines to help forecasters in the design and implementation of the regional prediction models that will be needed in the future.

In the following chapters, we will address the problem of regional power forecasting and upscaling from a statistical perspective. We will first characterize regional production and the relation existing between it and NWP variables.

We will then investigate different solutions for the combination of statistical single wind farm forecasting models to provide regional forecasts. The properties of different combinations schemes will be studied as well as the influence of different base-line models on forecast accuracy. From this guidelines will be put forward regarding the choice of base-line model and combination scheme.

Further, since explanatory variables (especially NWP forecasts) are considered as one of the primary sources of error in single wind farm forecasting, we will examine their influence in the regional setting. A framework for this examination will be proposed and the impact of forecast and measured variables will be examined. Then, given the large number of explanatory variables available in the frame of regional wind power forecasting, input variable selection will be addressed. The characteristics of explanatory variable selection in this setting will be exposed and selection methods that exploit these characteristics will be proposed.

CHAPTER 3

Characterization of regional forecasting

Abstract

The aim of this chapter is to propose a framework for the characterization and analysis of the regional wind power production and its relation to available explanatory variables. This framework considers regional wind power production both as a single time series and as an aggregate of multiple time series and allows examining its properties from both perspectives. The statistical smoothing of regional production is discussed in detail. Then, the relation between regional production and available NWP variables is characterized using a measure of the mutual information that exists between the different variables. This framework is then applied to characterize two case studies: one comprising 11 Irish wind farms and one considering the total regional production of the Jutland peninsula and of Funen island in Denmark. The results of the characterization of both cases are analyzed and compared to derive general conclusions on the most salient features of the regional wind power production.

3.1 Introduction

BEFORE implementing a forecasting model several questions must be answered concerning the forecasting problem at hand. Indeed, using an "off-the-shelf" model without carefully examining the properties of the problem is likely to lead to low quality forecasts. It is of prime importance to define as accurately as possible the aim of the forecasting model, in other words, what the needs of the forecast end-users are. Once this has been specified in detail, the second step is to determine the characteristics of the time series to be forecast. In depth knowledge of the nature of the time series permits to choose the characteristics a prediction model should have to adequately perform its task. Further, in cases where explanatory variables other then those derived from the target time series are available, analyzing the relation between these variables and the target time series is necessary to ensure correct use of the explanatory variables.

In this chapter, we first propose a definition of the regional forecasting problem. Here the aim is not to define the end-user needs as these may vary significantly from case to case. The purpose of this definition is to provide a general description of the common traits of regional forecasting found in the literature. Based on this definition, we then propose a characterization framework to investigate the properties of the regional forecasting problem.

The characterization of regional forecasting described here aims at analyzing the problem from three basic perspectives. We start by considering regional wind power production as a single time series. From this point of view, we use standard time series analysis tools to determine certain characteristics of regional production. Such results permit to determine necessary forecasting model properties. This analysis is strongly related to the type of analysis described by Box and Jenkins [30] and commonly carried out when ARMA type models are to be used.

After considering regional production as a single time series, we examine it as the aggregate of single wind farm production time series. Here, special emphasis is given to the evaluation of the statistical smoothing that takes place when single wind farm productions are aggregated into regional production. The aim of the investigation is to establish the influence of different parameters such as region size and wind farm dispersion on the amount of smoothing that takes place. More specifically, we investigate the influence of these parameters on the normalized variance of the regional production. We also consider the single wind farm productions as possible explanatory variables to be used by the forecasting model. We thus conduct an examination of the nature of the relation between single wind farm production and regional production. We examine this relation from a static and a dynamic perspective.

The last aspect we investigate is the relation between regional production and NWP forecasts. NWP forecasts are routinely used to enhance the accuracy of single farm production forecasting models for horizons typically greater than six hours ahead. Because these forecasts are sometimes available for some wind farms, using them as explanatory variables to enhance regional prediction accuracy is a logical step. Here we investigate the nature and strength of the statistical dependence between NWP forecasts and regional production. We particularly examine dependence differences in relation to the location for which the NWP data are provided. We also investigate the possibility of using average NWP forecasts as input to the forecasting models.

The chapter is organized in three parts. We begin by describing the tools that will allow us to determine the characteristics of regional forecasting. We then introduce two case studies on which the characterization is performed. One case study is located in Ireland and the other concerns the regional production of the Jutland peninsula and Funen Island in Denmark. In the third part of the chapter, the characterization results for both case studies are presented and analyzed.

3.2 Defining regional wind power forecasting

Regional forecasting is a term commonly used in the wind power forecasting field but there is no standard definition. When looking at regional wind power forecasting literature, the size of the regions considered vary widely mostly based on the case studies available to the authors for their research. In this sense, authors speak of regional wind power production for areas a few hundreds of kilometres in diameter (Ireland and Denmark) [38, 87], while others consider regional production for much larger regions (Germany, all of Europe) [24, 94].

What consensus can be extracted from the existing literature is that regional wind power is concerned with the aggregated production of wind farms spread over a certain area. This however is still a relatively loose definition since even the concept of wind farms can vary widely from country to country. For example, in Denmark and Germany some "wind farms" are actually single turbines owned by individuals or small cooperatives. In countries where wind power development has been driven by larger companies, wind farms usually comprise several wind turbines and have a total installed capacity of several tens of MWs. From the wind power forecasting literature, the most accepted or suggested definition is that a wind farm is comprised of a number of wind turbines sited in the same location and whose aggregated power output is to be predicted. Here, the wind farm can be considered as a group of wind turbines connected to the main grid through a single connection point. From this, regional power forecasting can be relatively defined as predicting the aggregated power output of wind turbines spread over a large area and for which the total production can be known by summing the productions of localized subsets of wind turbines.

This definition of regional power forecasting is necessarily relative with respect to scale. Indeed, this definition could be applied to single wind farm forecasting if the individual wind turbine productions of a wind park were known and the value to be forecast was the aggregated output of said turbines.

Moreover, recent on-shore wind farms have installed capacities of more than one hundred megawatts and offshore projects are currently being announced in the gigawatt range. Since a minimum distance between wind turbines (typically 4 or 5 rotor diameters) must be respected to minimize wake losses, such projects cover large areas. An example of such a project is the 1000 MW London Array offshore project [95], which is expected to occupy an area up to 245 square kilometres. If such projects are considered as wind farms, the forecasting techniques, which will be used to forecast their production, will likely be similar to those employed for regional wind power forecasting.

Another factor that can be taken into account when considering the difference between single wind farm forecasting and regional forecasting can be the hypotheses being made. Most often, a wind farm's production is considered as a single time series both in single wind farm forecasting and regional forecasting. Hence, the statistical smoothing of the productions of the wind turbines within a farm are neglected, the process being modelled in such cases is the relation between the explanatory variables and the wind farm's production. In other words, the wind farm's power curve is being modelled. However, some prediction models can consider each wind turbine individually and thus take into account the relative independence that exists between wind turbine productions much as a regional forecasting model considers the wind farm's productions to be relatively independent. This can be illustrated to some extent by the number of NWP grid points considered to describe the future weather situation over a wind farm. In most cases, wind farms occupy less area than that of the NWP model grid cells. Hence, the NWP data usually provided for single wind farm forecasting does not exceed five grid points (one grid point "inside" the wind farm and four grid points surrounding the wind farm). In many cases, a single interpolated value is provided. As wind farms become more widespread and the resolution of operational NWP models increases, the number of NWP grid points located within a wind farm's area will increase. In this case, the forecasting problem will be very similar to that of regional forecasting albeit on a smaller scale.

3.3 Time Series Analysis of regional production

In this section, the regional production is considered as a single time series. As such, we consider the available data as a set of observations measured sequentially through time. Due to the nature of data acquisition systems, the regional power output is measured at a discreet set of equally spaced time points, hence the available data are *discreet* time series which result from the sampling of *continuous* series. More over, the power data usually considered in wind power forecasting is averaged from data sampled at higher frequencies. A very common example is the averaging of ten-minute data to obtain one-hour measurements. Although higher frequency data has been used to increase forecast accuracy in wind forecasting [96], this aspect is outside of the scope of the study conducted here.

The first aspect we will investigate is the stationarity of regional production. Here the production can be seen as a stochastic process. A stochastic process can be defined as a family of random variables X_t indexed by time $t \in T$. The time series is a finite realization of the process, where x_t is an observation at time t of the random variable X_t. The stochastic process is said to be *strict-sense stationary* if its statistical properties are invariant to a shift of the origin, that is X_t and X_{t+k} have the same statistics for any k [97]. Hence, a series can be said to be stationary if its statistical properties are independent of the time during which it is observed [40]. Thus, we can detect non-stationarity by examining successive moving averages $M_d(t)$ of sufficiently high order d applied to the time series. The moving average is defined as:

$$M_d(t) = \frac{1}{d} \sum_{i=t-\frac{(d-1)}{2}}^{t+\frac{(d-1)}{2}} x_i \tag{3.1}$$

If the series is stationary the moving averages should only vary slightly, or not at all, if d is sufficiently large. If the moving average does vary, then the series is non-stationary. Another criterion that can establish the non-stationarity is the moving variance defined as:

$$V_d(t) = \frac{1}{d-1} \sum_{i=t-\frac{(d-1)}{2}}^{t+\frac{(d-1)}{2}} (x_i - \bar{x})^2 \tag{3.2}$$

where $\bar{x}$ denotes the mean of all observations. As with the moving average, the moving variance should be constant for sufficiently large order d. Higher moving moments can be computed although this is seldom encountered in the literature; the moving average and moving variance are usually sufficient to determine a time series' non-stationarity.

Another aspect which can be examined is the autocorrelation r_k of the time series. If we note the time series by $x_1, x_2, \ldots, x_N$, the sample autocovariance is defined as [33]:

$$c_k = \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x}) \tag{3.3}$$

for $k = 0, 1, 2, \ldots$, and the sample autocorrelation at lag k is defined as:

$$r_k = \frac{c_k}{c_0} \tag{3.4}$$

By examining the sample autocorrelations obtained for successive lags, insight into the behaviour of the time series can be gained. For example, if r_1 is high, then there is a strong linear relation between x_t and x_{t-1}. In essence, the autocorrelation provides information on the influence of the past values of the time series on the current value. If the autocorrelation falls to zero after the first few time lags, then the process is likely to by influenced only by recent phenomena. If however, significant autocorrelation exists for large lags, then the time series is likely non-stationary [33]. Further, significant autocorrelations present at equally spaced time lags, indicate the presence of seasonality in the time series. This implies that the values of the time series are not only dependent on time, but that a regularly recurring phenomenon influences the values of the series. In order to establish the significance of the autocorrelation, the standard error of r_k can be calculated by [40]:

$$se_{r_k} = \frac{1}{\sqrt{N}} \tag{3.5}$$

The sample autocorrelation coefficients of random data are asymptotically normally distributed with mean $1/N$, and standard deviation $1/\sqrt{N}$ [40]. From this, a significance test can be deduced. For example, 95 percent of all sample autocorrelation coefficients must lie within plus or minus 1.96 standard errors. Hence, if:

$$-1.96 se_{r_k} \leq r_k \leq 1.96 se_{r_k} \tag{3.6}$$

then the time series can be concluded to be random. If the autocorrelation lies outside this range, then it can be concluded to be significant and not due to sampling variations.

3.4 Regional production as the sum of individual productions

3.4.1 Statistical smoothing of regional production

The focal point of this section is the variance of the aggregated production of several wind farms. Indeed, as demonstrated in [37], the forecast error of the climatological mean model sets the upper bound on the error of wind power forecasting models. The mean square error for the climatological mean model is :

$$\begin{aligned} E\left((P - E(P))^2\right) &= E(P^2) - 2E\left(PE(P)\right) + E(P)^2 \\ &= E(P^2) - E(P)^2 \\ &= V(P) \end{aligned} \tag{3.7}$$

where P is the wind power production time series, $E(\cdot)$ denotes the expected value, $E(P)$ is the expected value of the production, also known as the climatological mean, and $V(P)$ denotes the variance of the production. The variance of the wind power series can thus be considered as an upper bound of the prediction error. Therefore, investigating how the variance of the aggregated production behaves with respect to the properties of the individual wind farm productions is of interest. An important point is the comparison of the variance of the aggregated power with respect to the variance of the individual wind farm productions. In the following paragraphs, we review previously published studies on this subject. Following this, we define a smoothing factor and an experimental framework that allows us to characterize aggregated wind farm productions and thus complete existing results.

The geographical distribution of wind farms can lead to the statistical smoothing of the power production. When considering the sum of wind farm productions, individual variations cancel each other to a certain extent. This phenomenon has been reported in many studies aiming at estimating the impact of large-scale wind generation on electricity grids [94, 98]. Most often, such results can be found in studies of the potential capacity credit of wind power. In some cases, the regional wind power production is simulated from wind speed readings measured at weather stations. This is the case in [94, 98], where the authors found that regional production presents slower variations than single wind farm production. In both references, the normalized variance of the difference between lagged time series was shown to be smaller for regional production than for single wind farm production. Also, extreme production events, i.e.: no production and full production, occur with much less frequency than in the case of single wind farms. This can be explained by the de-correlation of wind conditions when considering the aggregated production over large areas. In such cases there will almost always be some wind somewhere, whereas having high wind speeds everywhere becomes very unlikely.

In [94] the relation between the variance of aggregated wind farm production and the size of the region over which the farms were scattered was experimentally examined. In this study, wind farm production was simulated by applying an average power curve (representative of common wind turbines) to the wind speed measurements from 60 meteorological stations spread over Europe. The results of this study show that the correlation between the productions of two wind farms is inversely proportional to their distance. It was found that null and negative correlations consistently appear for distances superior to 1500 kilometres. An exponential function was proposed to model the relation between correlation and distance:

$$Corr(P_i^{wf}, P_j^{wf}) = \exp(\frac{-d_{i,j}}{a}) \tag{3.8}$$

where $Corr(P_i^{wf}, P_j^{wf})$ is the correlation between the productions of wind farms wf_i and wf_j, $d_{i,j}$ is the distance between the wind farms, and a is a fit parameter. It was also found that the normalized standard deviation of aggregated wind farm productions decreases when the size of the region in which the farms are located increases. In this study, the region size was defined as the maximum distance of any one farm to the centre of gravity of the wind farm coordinates. Another conclusion from this study was that the variation of the normalized standard deviation decreases with the number of wind farms included in the aggregation.

The results of this study are important since they allow us to state that for this study the variance and therefore the upper bound of the forecast error decreases with region size and that the number of wind farms considered in an aggregation influences the likelihood of this reduction. Although the variance of a time series does not directly measure its "predictability", lower variance can indicate higher predictability. Therefore, the forecast error obtained when predicting the aggregated production of several wind farms can be expected to be lower than that observed when predicting individual productions. However, the author in [94] did not investigate the normalized variance of aggregated productions with respect to the normalized variance of the individual productions.

The comparison of aggregated variance and individual variance was indirectly examined in [24] where the authors investigated the standard deviation of regional forecast error and the standard deviation of single wind farm forecast error. In this study on German wind farms, they showed that the correlation between the forecast errors obtained for two wind farms is inversely proportional to the distance between the wind farms. In the same spirit as [94] they propose to model this relation with piece-wise exponential functions such as:

$$Corr(e_i^{wf}, e_j^{wf}) = a\exp(\frac{-d_{i,j}}{b}) \tag{3.9}$$

where $Corr(e_i^{wf}, e_j^{wf})$ is the correlation between the forecast error obtained for wind farms wf_i and wf_j, $d_{i,j}$ is the distance between the wind farms, and a, b are fit parameters. They then defined smoothing as the ratio between the normalized standard deviation of the forecast error for the aggregated production —$\sigma_{ensemble}$— and the mean normalized standard deviation of the error computed at the individual sites: σ_{single}. They found that this ratio decreases with increasing region size (in this case the radius of the circle containing the wind farms) for the case of Germany. In order to generalize their results and examine the influence of region size and the number of farms considered in an aggregation on the smoothing of the errors, the authors proposed a numerical simulation of the smoothing effect. They modelled the variance of the aggregated production forecast error by using the formula for the variance of a sum of random variables (Equation 3.10)

$$V(\sum_{i=1}^{N} e_i^{wf}) = \sum_{i,j} \sigma(e_i^{wf})\sigma(e_j^{wf})Corr(e_i^{wf}, e_j^{wf}) \tag{3.10}$$

in which they replaced $Corr(e_i^{wf}, e_j^{wf})$ with the function of distance defined in Equation 3.9. By taking the hypothesis that for all wind farms $\sigma(e_i^{wf}) = 1$ and by randomly drawing the coordinates of the wind farms, they were able to determine the standard deviation of aggregated production forecasts. By repeating this simulation a number of times they showed that the expected value of $\frac{\sigma_{ensemble}}{\sigma_{single}}$ decreases with region size and that for a given region size it decreases and attains a limit when the number of wind farms considered in the region increases.

The results in this study are important for two main reasons. First, they introduce the concept of smoothing as being the relation between the normalized variances of the aggregated series and the individual series. Secondly, these results corroborate the idea that the reduction on the upper bound of the error derived from aggregating single wind farm productions does lead to an actual reduction of the regional forecast error obtained using advanced forecasting models. However, the results presented in this study only considered the expected value of the reduction of the error in a hypothetical setting but did not examine the values that can be obtained when aggregating different numbers of real wind farms in regions of varying size.

In the following paragraphs, we propose to show that with the hypotheses made in [24], the experimental results obtained in that study can be derived analytically for the normalized variance of both the aggregated production and the forecast error. We then propose a smoothing factor based on the idea of comparing aggregated production variances and single wind farm variances. We then define an experimental framework to investigate this factor's behaviour with respect to region size, the number of wind farms in an aggregation and the mean distance between wind farms.

As shown in [24, 94] the correlation between single wind farm productions and single wind farm power forecast errors is inversely proportional to the distance separating the wind farms. Therefore, the model proposed in [24] can be used to estimate the variance of the aggregated time series. The model can be expressed as follows:

$$V(\sum_{i=1}^{N} X_i^{wf}) = \sum_{i,j} \sigma(X_i^{wf})\sigma(X_j^{wf})C(d_{i,j}) \tag{3.11}$$

where $C(d_{i,j})$ models the correlation between the series X_i^{wf} and X_j^{wf} (associated to wind farms wf_i and wf_j) as a function of the distance $d_{i,j}$ between them. We now make the same hypotheses as in [24]:

Hypothesis 1: Let $d_{i,j}$ be a random variable whose expected value $E(d_{i,j})$ is an increasing function of the region size R. In this case $C(d_{i,j})$ is also a random variable.

Hypothesis 2: Let $C(d_{i,j})$ be independent of the number of wind farms in the region and

$$\lim_{R\to\infty} E\left(C(d_{i,j})\right) = 0. \tag{3.12}$$

Hypothesis 3: For all wind farms wf_i in a region, let $\sigma(X_i^{wf}) = \sigma$ and the installed capacity of each wind farm $P_{inst,i}^{wf} = P_{inst}$.

This is a generalization of the hypothesis taken in [24] where for all wind farms $\sigma(e_i^{wf}) = 1$.

By considering Hypothesis 3 and Equation 3.11 the normalized variance of the aggregated series can be written:

$$\frac{V(\sum_{i=1}^{N} X_i^{wf})}{(NP_{inst})^2} = \frac{\sigma^2}{(NP_{inst})^2}(N + 2\sum_{i,j:i<j} C(d_{i,j})). \tag{3.13}$$

The expected value of the normalized variance can then be derived from Hypothesis 1 and Equation 3.13:

$$\begin{aligned} E(\frac{V(\sum_{i=1}^{N} X_i^{wf})}{(NP_{inst})^2}) &= \frac{\sigma^2}{(NP_{inst})^2}(N + 2\sum_{i,j:i<j} E(C(d_{i,j})) \\ &= \frac{\sigma^2}{(NP_{inst})^2}(N + (N^2 - N)E(C(d_{i,j})) \\ &= \frac{\sigma^2(1 - E(C(d_{i,j})))}{NP_{inst}^2} + \frac{\sigma^2 E(C(d_{i,j}))}{P_{inst}^2}. \end{aligned} \tag{3.14}$$

From Equation 3.14 and Hypothesis 2 the limit of the expected value of the normalized variance when the number of wind farms N and the region size R go to infinity can be derived. When the number of wind farms increases, the expected value of the normalized variance will decrease and attain a limit:

$$\lim_{N\to\infty} E(\frac{V(\sum_{i=1}^{N} X_i^{wf})}{(NP_{inst})^2}) = \frac{\sigma^2 E(C(d_{i,j}))}{P_{inst}^2} \tag{3.15}$$

From this we can conclude that for a region of given size and expected correlation, increasing the number of wind farms will lead to an increase in the probability of witnessing a reduction of the normalized variance of the aggregated series. Also, given the derivative of the expected value of the normalized variance with respect to N, values close to the limit will be reached once a few farms have been considered in the region. With respect to region size, the limit of the expected value of the normalized variance when R increases is:

$$\lim_{R\to\infty} E(\frac{V(\sum_{i=1}^{N} X_i^{wf})}{(NP_{inst})^2}) = \frac{\sigma^2}{NP_{inst}^2} \tag{3.16}$$

From this limit, it becomes clear that the expected value of the normalized variance of the aggregated series will decrease with the size of the considered region.

These results corroborate the experimental results published in [94] for the normalized variance of aggregated wind farm productions where the variance of the aggregated production was shown to decrease with region size and the variation of observed variance values was shown to decrease with the number of wind farms considered. They also confirm the numerical results published in [24]. With respect to the smoothing of the error, when the aggregated variance is divided by the mean normalized variance, the limits derived in Equation 3.15 and Equation 3.16 become:

$$\lim_{N\to\infty} E(\frac{P_{inst}^2 V(\sum_{i=1}^N e_i^{wf})}{\sigma^2(NP_{inst})^2}) = E(C(d_{i,j})) \tag{3.17}$$

$$\lim_{R\to\infty} E(\frac{P_{inst}^2 V(\sum_{i=1}^N e_i^{wf})}{\sigma^2(NP_{inst})^2}) = \frac{1}{N}. \tag{3.18}$$

These limits are also valid for the ratio between the aggregated variance and the mean variance of the production time series. However, it must be stressed that these results concern the expected value of the variance, but give no indication on the values that can be observed. For this reason, we propose a methodology to examine the behaviour of the smoothing that takes place in real cases.

As suggested in [24], the smoothing effect is a relative phenomenon where the behaviour of aggregated time series is compared to the average behaviour of the individual time series. With respect to wind power production time series, it can be characterized by examining the variance of the production of a single wind farm and the variance of regional production. The simplest way to put this phenomenon into evidence is to compare the normalized variance of regional production to the average normalized variance of the productions of the wind farms in the region. Therefore, smoothing of the production can be said to occur if the following inequality is verified:

$$\frac{V(\sum_{i=1}^N P_i^{wf})}{(P_{inst}^{reg})^2} \leq \frac{1}{N}\sum_{i=1}^{n} \frac{V(P_i^{wf})}{\left(P_{inst,i}^{wf}\right)^2} \tag{3.19}$$

where P_i^{wf} and $P_{inst,i}^{wf}$ respectively denote the production and installed capacity of wind farm wf_i, $P_{inst}^{reg} = \sum_{i=1}^N P_{inst,i}^{wf}$ is the installed capacity in the region and N the number of wind farms in the region.

Given the properties of the variance, Equation 3.19 can be written:

$$\frac{1}{(P_{inst}^{reg})^2}\left(\sum_{i=1}^N V(P_i^{wf}) + 2\sum_{i,j:i<j} cov(P_i^{wf}, P_j^{wf})\right) \leq \frac{1}{N}\sum_{i=1}^{N} \frac{V(P_i^{wf})}{\left(P_{inst,i}^{wf}\right)^2} \tag{3.20}$$

where $cov(P_i^{wf}, P_j^{wf})$ denotes the covariance between the productions of wind farms wf_i and wf_j.

TABLE 3.1: *Wind farm combinations for all cardinalities, in the case of four wind farms. Wind farms are referred to by number.*

Combination cardinality	Combinations					
2	1-2	1-3	1-4	2-3	2-4	3-4
3	1-2-3	1-2-4	1-3-4	2-3-4		
4	1-2-3-4					

From Equation 3.19 it is clear that when the productions of the wind farms are independent, i.e. their covariance is null, the smoothing is likely to be more important than when the productions are not independent, i.e. non null covariance. This is likely because the existence of negative covariance cannot be dismissed as impossible. Having negative covariance would mean that the productions of certain wind farms simultaneously vary in opposite directions. Summing productions that vary in opposite directions can lead to very stable production and to a lower variance of the total production than that of independent productions.

In order to quantify the amount of smoothing that occurs when aggregating N wind farms we consider the following smoothing factor S:

$$S = 1 - \frac{\frac{V(\sum_{i=1}^{N} P_i^{wf})}{\left(\sum_{i=1}^{N} P_{inst,i}^{wf}\right)^2} -}{\frac{1}{N}\sum_{i=1}^{n} \frac{V(P_i^{wf})}{\left(P_{inst,i}^{wf}\right)^2}} \tag{3.21}$$

The smoothing factor S quantifies the reduction of production variance observed when comparing the aggregated production variance to the average production variance of the wind farm within the aggregation.

In an experimental setting where production time series are available for a number of wind farms, the behaviour of the smoothing factor can be examined for all possible wind farm combinations, see Table 3.1. Each wind farm combination can then be considered as a “virtual region”. To investigate the relation between the smoothing factor and region size, the virtual region’s size can be defined as the largest distance between the wind farms that belong to it. Since wind farm distribution plays an important role in the likely observed smoothing factor, we propose to assess the distribution of the wind farms within each virtual region by comparing the average distance between wind farms to the region size. If the average distance is small compared to the region size, then many farms are concentrated in one area and smoothing is likely to be lower than for other similarly sized virtual regions.

By applying this methodology to different case studies, the relation between the smoothing factor and the region size, the wind farm concentration and the number of wind farms within each region, can be examined. Average trends such as those put forward in [24, 94] can be further confirmed, but individual cases of interest can also be highlighted.

3.4.2 Statistical relation between regional production and single wind farm productions

Today, most modern wind farms are equipped with SCADA systems capable of measuring and transmitting production data. These on-line production measurements therefore provide very useful information to wind power forecasting models. If all wind farms within a region are with SCADA systems computing the total regional production is trivial. In fact, having this data can allow implementing simple models such as Persistence or more advanced time series forecasting models.

This ideal case is however seldom encountered in real-world regional forecasting cases. Very often, only a fraction of the wind parks installed in a region can provide on-line production data. This limit can stem either from historical reasons: older wind farms are not always equipped with data transmission equipment; or from more practical reasons: the party interested in obtaining regional forecasts might not have direct access to certain wind farm production measurements[1]. When on-line measurements of the regional production are unavailable, the forecasting models have to extrapolate from whatever on-line data is available. Because of this, examining the nature of the relation that exists between the production of a single wind farm, or a group of wind farms, and the regional production is necessary.

One possibility to investigate this relation is to use the correlation coefficient $\rho_{X,Y}$ which is defined as:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E((X-\mu_X)(Y-\mu_Y))}{\sigma_X \sigma_Y} \tag{3.22}$$

where X and Y are two random variables with expected values μ_X and μ_Y and standard deviations σ_X and σ_Y. The correlation coefficient measures the linear dependence between two variables. If two variables are independent then their correlation will be zero, and if a perfect linear relation exists between both variables, the correlation will be 1 in the case of a positive relation, or -1 in case of a negative relation. For values between 1 and -1 the correlation measures the strength of the relation. The sample correlation of X and Y can be estimated using the Pearson correlation coefficient defined as:

$$r_{xy} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} \tag{3.23}$$

where $\bar{x}$ and $\bar{y}$ are the sample means of x_i and y_i, s_x and s_y are the sample standard deviations of x_i and y_i, and N the number of available observations in the dataset.

The correlation coefficient will be used here to measure the linear relation between single wind farm productions and regional productions. The differences in correlation between different wind farms can help detect differences in the contribution of each wind farm to the total regional generation. The correlation coefficient can also be applied to

[1] It can be noted that grid codes are being modified to ensure that SCADA data is provided to the TSO at least for wind farms whose installed capacity is above a certain threshold.

quantify the relations between the aggregated production of a sub-group of wind farms and the regional production. In this way the representativity of the production of different subssets of wind farms can be estimated and the relative importance of possessing production measurements for all wind farms in a region can be estimated.

The correlation coefficient can also be computed in a sliding window approach similar to the moving average and moving variance described in section 3.3. In this way, the evolution of the correlation coefficient between two productions can be observed. As with the moving average, this can highlight non-stationarity in the linear relation between a wind farm's production and the total regional production. By examining the evolution of the moving correlation and other moving statistics, different phenomena, which might need to be considered when building a forecasting model, can possibly be revealed. This can be particularly important in upscaling cases where the on-line production measurements do not provide complete information on the total production.

3.4.3 Statistical relation between regional production and numerical weather predictions

Aside from production measurements, the most common explanatory data used today to compute wind power forecasts are NWP forecasts. As mentioned before, using these forecasts greatly increases the accuracy of wind power forecasts especially for horizons greater than 6 hours ahead. This is also true in the case of regional forecasting. Another property of NWPs, which could be exploited in the regional forecasting case, is the fact that they are computed on a grid spanning areas typically greater than the region for which power forecasts are being computed. Because of this, NWP forecasts can provide information for wind farms for which no power measurements are available on-line.

In order to examine how well NWP forecasts can explain regional production, using the correlation coefficient is not the most appropriate method. As mentioned in subsection 3.4.2, the correlation coefficient and Pearson's sample correlation coefficient measure the strength of linear relations between variables. However, the relation between wind farm production and relevant atmospheric variables not always linear. This is clearly the case for wind speed, which has a cubic relation to the power output of a wind turbine (see Equation 2.1). In order to measure non-linear relations, in essence the level of dependence of the variables, we choose to use mutual information instead of correlation.

The average mutual information $I(X;Y)$ between two random variables X and Y can be defined as [99]:

$$I(X;Y) = D(P_{XY} \| P_X \times P_Y) \tag{3.24}$$

where $D(P_{XY} \| P_X \times P_Y)$ denotes the divergence of the probability distribution P_{XY} with respect to the product of the marginal distributions P_X and P_Y of X and Y.

The divergence, also called the relative entropy or Kullback-Leibler divergence, can be defined for a discrete random variable X as:

$$D(P_X \| M_X) = \sum_{\omega \in \Omega} P_X(\omega) \ln \frac{P_X(\omega)}{M_X(\omega)} = \sum_{\omega \in \Omega} p(\omega) \ln \frac{p(\omega)}{m(\omega)} \tag{3.25}$$

where P and M are two measures defined on the measurable space (Ω, F), P_X and M_X the two induced distributions of X, and p and m the corresponding probability mass functions (i.e. $p(\omega) = P_X(\omega) = P(X = \omega)$). When the random variable X is continuous, the sum in Equation 3.25 becomes an integral:

$$D(P_X \| M_X) = \int f(\omega) \ln \frac{f(\omega)}{g(\omega)} \, d\omega \tag{3.26}$$

where f and g denote respectively the probability density functions induced by P_X and M_X.

From this, the average mutual information between two discreet random variables X and Y defined on the same probability space (Ω, F, P) is defined in terms of the joint distribution P_{XY} and the marginal distributions P_X and P_Y as:

$$I(X;Y) = \sum_{x,y} P_{XY}(x,y) \ln \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)} = \sum_{x,y} p_{XY}(x,y) \ln \frac{p_{XY}(x,y)}{p_X(x)p_Y(y)} \tag{3.27}$$

where p_{XY} is the joint probability mass function for (X,Y) and p_X and p_Y are respectively the probability mass functions for X and Y. Similarly, for two continuous random variables, the average mutual information can be defined as:

$$I(X;Y) = \int f_{XY}(x,y) \ln \frac{f_{XY}(x,y)}{f_X(x)f_Y(y)} \, dxdy \tag{3.28}$$

where f_{XY} is the joint probability density function for (X,Y) and f_X and f_Y are respectively the marginal probability density functions for X and Y.

Given two random variables X and Y, then $I(X;Y) \geq 0$, and $I(X;Y) = 0$ if and only if X and Y are independent.

In a sense, mutual information quantifies the "distance" between the joint distribution of X and Y and what the joint distribution would be if X and Y were independent. In a more intuitive sense, mutual information quantifies the amount of information shared by X and Y. In other words, it quantifies how much knowing the outcome of one variable reduces the uncertainty on the outcome of the other. If the variables are independent, then, knowing the outcome of one variable does not reduce the uncertainty concerning the outcome of the other, in this case $I(X,Y) = 0$. If, however, both variables are identical, then knowing the outcome of one variable equates to knowing the outcome of the other, in this case $I(X,Y) = I(X,X) = H(X)$ where $H(X)$ denotes the information entropy of X. The information entropy $H(X)$ can be seen as the uncertainty on the outcome of X.

Computing the average mutual information requires estimating the probability mass functions or the probability density functions of the variables under consideration. We consider here the meteorological variables and the power production as continuous random variables. To estimate the probability density functions, two broad classes of methods are applicable: parametric methods and non-parametric methods. In parametric methods, given a parametric density family $f(\cdot|\theta)$, the aim is to obtain the best estimator $\hat{\theta}$ of θ. Non-parametric methods aim at obtaining an estimate $\hat{f}$ of the density $f(\cdot)$. Parametric estimation involves two phases: *specification* and *estimation*. Specification deals with the selection of the density family to be used. Estimation deals with the estimation of the density parameters. Although many parameter estimation methods are available for different density families, the application of such methods requires that a hypothesis be made on the density that underlies the data on which the fitting is being carried out. In our case, the underlying densities are not necessarily known *a priori*. This can lead to inaccurate estimation of the probability densities and hence to inaccurate mutual information estimates. For this reason, parametric methods will not be used in the computation of mutual information. Here, we will use kernel density estimation, a non-parametric method well suited for mutual information estimation [100, 101].

From a practical point of view, the average mutual information will allow to compare different explanatory variables as they relate to the total production. For example if $\hat{W}s(k)$ is the predicted wind speed for horizon k, $\hat{W}d(k)$ is the predicted wind direction for k and $P^{wf}(k)$ the measured production of a wind farm at time k, then if $I(\hat{W}s(k), P^{wf}(k)) > I(\hat{W}d(h), P^{wf}(k))$ this means that the dependence between wind speed forecasts and wind farm production is stronger than the dependence between wind direction forecasts and wind farm production. Hence, a forecasting model which uses only one explanatory variable will provide better production forecasts if it uses wind speed forecasts instead of wind direction forecasts as the explanatory variable.

3.5 Scope of the study

The different tools presented in the previous sections of this chapter are used here to characterize two case studies. The first case study consists of eleven wind farms located in Ireland; the second one consists of twenty-three wind farms located in Denmark.

3.5.1 Irish case-study

Ireland is a country with a high wind potential. Wind prediction systems, such as the one developed in the frame of the More-Care project [64], are currently used operationally by the system operator. Here, eleven wind farms corresponding to a total power of 70 MW are considered.

The geographical distribution of the eleven wind farms is uneven over the region. The majority of the wind farms are located in the northwest of the island. Figure 3.1 presents the approximate locations of the wind farms under consideration. Table 3.2 lists these farms and their relative installed capacity.

The available data consist of production measurements for each wind farm as well as NWP forecasts for each wind farm. The production measurements are hourly mean power

TABLE 3.2: *List of wind farms in the Irish case study.*

Wind farm number	Wind farm Name	Relative Nominal Power (% of total Nom. power: 70.3 MW)
1	Bellacorik	9.25
2	Cark	21.34
3	Corrie Mountain	7.11
4	Cronalaght	7.11
5	Crockahenny	7.11
6	Curabwee	6.68
7	Drumlough	7.11
8	Golagh	21.34
9	Inverrin	3.98
10	Kilronan	7.11
11	Spion Kop	1.85

outputs.

The NWP forecasts are provided by Met Éireann, the Irish meteorological service. The Hirlam model used to compute these forecasts has a spatial resolution of roughly 30 km. The NWP forecast runs are provided four times a day (at 00:00, 06:00, 12:00 and 18:00). The forecast horizon is forty-eight hours, with an hourly time resolution. The forecasts are provided as interpolated values for the sites of the wind farms. The available meteorological variables include wind speed and wind direction for altitudes: 10 meters a.g.l. and for the Hirlam model levels 22, 23 and 24. Temperature forecasts are also given for 2 meters a.g.l. and for the above model levels.

The production data as well as the NWP forecasts cover a period between 5 March 2001 (12h00) and 16 August 2002 (00h00).

3.5.2 Danish case-study

Denmark has the third largest installed wind capacity in Europe with 3.136 GW installed by the end of 2006 [102]. The case study available to us concerns 23 wind farms located in the Jutland peninsula and Funen Island. The total capacity of these wind farms is 192.15 MW. The total installed capacity of 2.2 GW of the Jutland-Funen is also considered.

The geographical distribution of the available wind farms is relatively even over the region. Figure 3.2 presents the locations of the wind farms under consideration. Table 3.3 lists the 23 wind farms and their installed capacity normalized with respect to the sum of their nominal powers.

The production data consist of measured mean hourly power output for the 23 wind farms as well as for the total power output of all the wind farms located in the Jutland-Funen area. The NWP data for the 23 wind farms are also at our disposal. The measured and NWP data span 1 year and 7 months from 1 January 2003 to 31 July 2004.

The NWP forecasts are computed using the Hirlam model, and are provided by DMI,

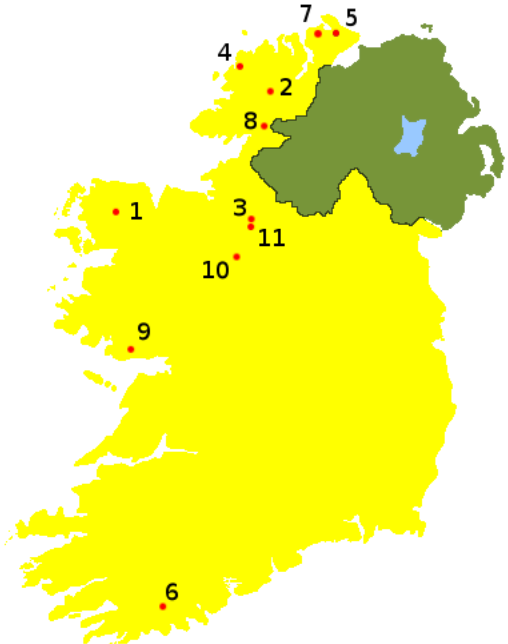


FIGURE 3.1: *Map of Ireland showing the approximate locations of the wind farms that constitute the Irish case study*

TABLE 3.3: *List of wind farms in the Danish case study. Nominal power is given as percentage of total nominal power: 192.15 MW.*

Wind farm number	Wind farm Name	Nominal Power (% Tot.)	Wind farm number	Wind farm Name	Nominal Power (% Tot.)
1	Abild	1.30	13	Klim	10.93
2	Arrild	3.43	14	Nørrekær Enge	8.93
3	Birkmose	1.77	15	Nordjyllandsværket	4.16
4	Brøns	1.67	16	Nr. Økse Sø	9.37
5	Dageløkke	0.16	17	Rejsby	12.18
6	Dræby	1.37	18	Ravning	0.83
7	Frederikshavn	5.52	19	Tjæreborg	3.90
8	Fjaldene	3.38	20	Torrild/Svinsager	0.23
9	Gl.Vrå Enge	5.07	21	Tunø Knob	2.60
10	Hanstholm	0.82	22	Vedersø Kær	7.81
11	Hollandsbjerg	8.59	23	Velling Mærsk	4.73
12	Højer/Emmerlev	1.25			

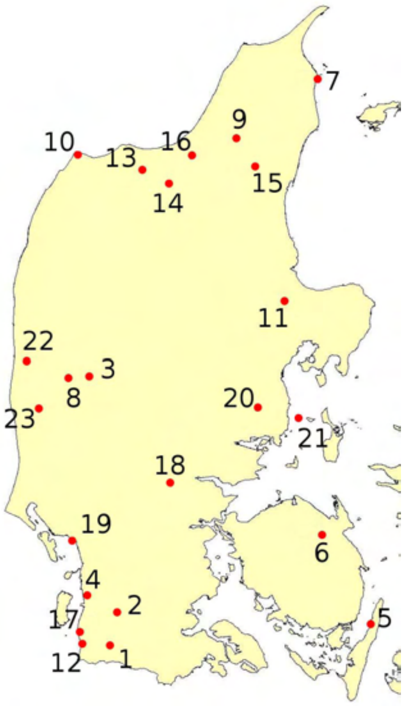


FIGURE 3.2: *Map of Jutland peninsula and Funen Island, showing the locations of the wind farms that constitute the Danish case study*

the Danish meteorological institute. The Hirlam model used here has a spatial resolution of 0.15°. The NWP forecast runs are provided twice a day (at 00:00 and 12:00). The forecast horizon is 48 hours, with an hourly time resolution. The forecasts are provided as interpolated values for the sites of the wind farms. The available meteorological variables include: wind speed and wind direction for: 10 meters a.g.l. and for Hirlam model levels 31, 38, 39 and 40.

This case study is very different from the Irish case study. In the Irish case, the total regional power output we consider is the sum of the 11 wind farms' production. In the present case, the 23 wind farms for which production data are available represent only 9% of the total installed capacity of the region.

3.6 Characterization of regional production

In this section, we use the tools described in previous sections of this chapter to examine the characteristics of regional production from a forecasting point of view. We will study regional production as a single time series and we will investigate its relation with the available explanatory variables.

3.6.1 Time series analysis of regional production

Moving averages as well as moving variances provide information on the stationarity of a time series. Moving averages can be computed for different orders or sliding window sizes. Choosing the order of a moving average is the result of a compromise. It depends of course on the size of the data set, a moving average cannot have an order larger than the dataset. The second aspect a moving average order must take into account is sufficient smoothing. Choosing too small a moving average will reveal "parasite" non-stationarity due to under smoothing. On the other hand, too large an order might smooth out all non-stationarity. For these reasons, we have chosen to examine the six-month moving averages of the regional production of both case studies. A six-month order, equivalent to 4393 hours, reasonably guaranties avoidance of "parasite" non-stationarity while revealing long-term trends.

In Figure 3.3, the six-month moving averages computed on the total power production of the Irish case study (70 MW) and the Danish case study (2.2 GW) are presented. The straight lines represent the global, or climatological, means for both case studies. It should be noted that although both moving average curves seem to follow the same trend they are not coincidental. From these results, it appears that both productions are clearly non-stationary. This is not a novel result, it is well known that seasonal production variations are common in Europe. There are typically higher wind speeds in winter and lower wind speeds in summer. However, the level of variation around the global mean is rather important. Furthermore, a difference in the amount of variation can be seen when the case studies are compared with each other. Independently of the respective load factors, the Irish case study presents higher variations around the global mean than the Danish case. This can possibly be linked to the load factor[2] of each case. Higher load factor regions are likely to present higher variations since low production periods will be compensated by high production periods, which will lead to important fluctuations around the mean production.

In Figure 3.4, the six-month moving variances computed on the total power production of the two case studies are presented. The straight lines represent the global variance for both case studies. Here again, it is clear that the variance is not invariant through time, both series are not second-order stationary. As with the moving averages, the Irish case presents stronger variability of its variance than the Danish one. An interesting phenomenon that must be noted, especially for the Danish production, is the fact that the variance seems to have a somewhat stepwise variation. In several places along the moving variance curves plateaus appear, noticeably around hour 3000, which seem to indicate a locally stable variance. In many places, one plateau is followed by another of different value after a rather sharp variation. The Irish case does not present such clear plateaus although the moving variance seems to vary less rapidly than the moving average. The appearance of plateaus in the Danish case and not in the Irish case might be linked to the difference in scale of both case studies.

Another way to detect non-stationarity is to examine the correlogram of the production

[2]The load factor of a wind park is defined as the ratio between a park's yearly production and its theoretical maximum yearly production. The latter corresponds to its nominal capacity multiplied by 8760, the number of hours in a year.

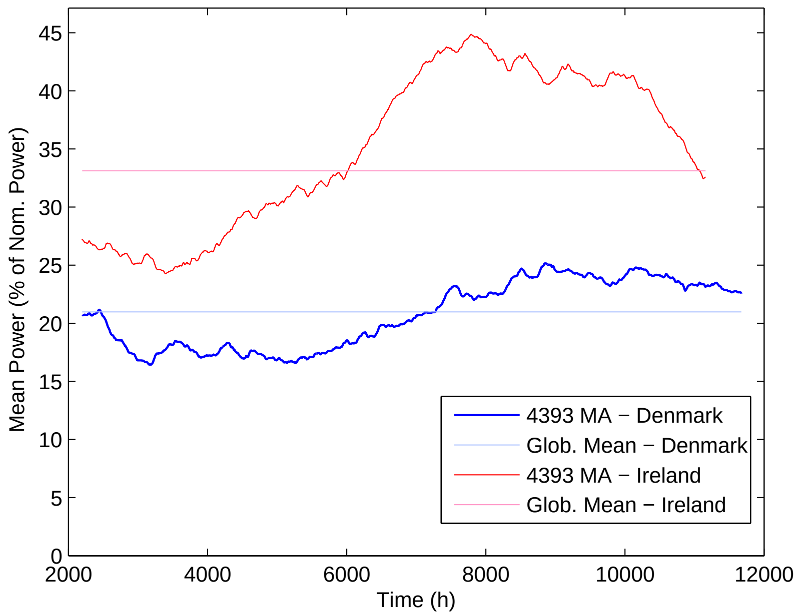


FIGURE 3.3: *Six month (4393 hours) moving average (MA) computed for the total production measurements for the Irish and Danish case studies. The moving averages are compared to the global, or climatological, mean of each case.*

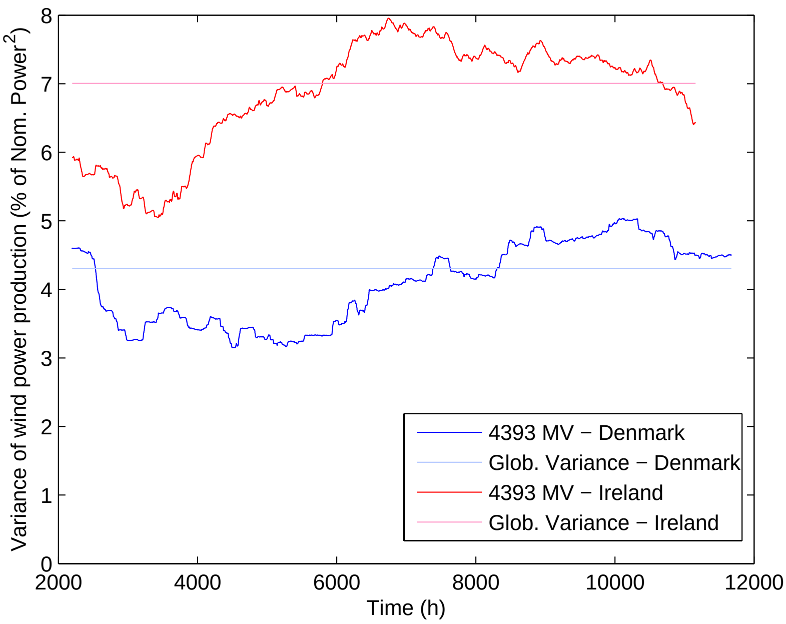


FIGURE 3.4: *Six month (4393 hours) moving variance (MV) computed for the total production measurements for the Irish and Danish case studies. The moving variances are compared to the global variances.*

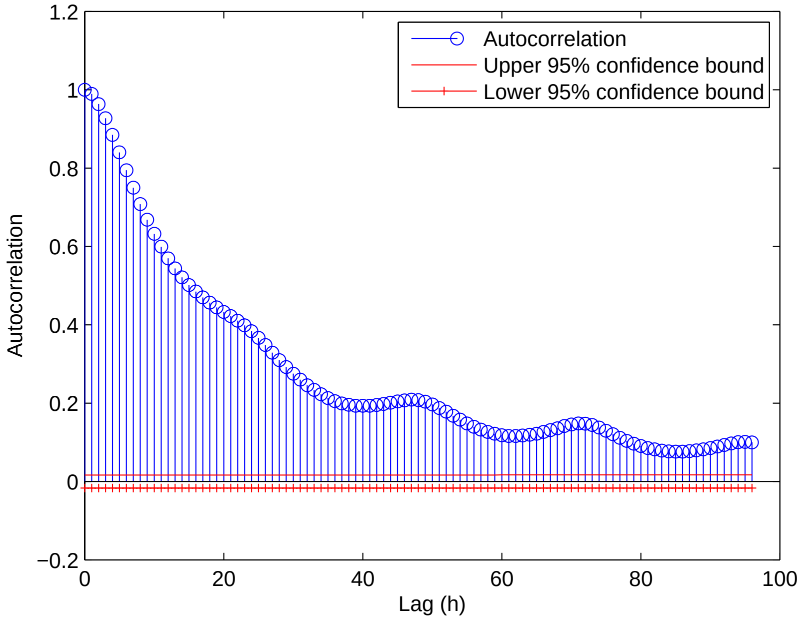


FIGURE 3.5: *Sample correlogram of the Jutland-Funen production time series. The correlations are computed up to a lag of 96 hours.*

time series. In Figure 3.5 the correlogram computed on the total Jutland-Funen production is presented. The most striking aspect in that figure is the fact that for all lags the autocorrelation remains significant; all the correlation values are well above the 95% significance interval. The second important aspect that the correlogram reveals is the high correlation values for the first time lags. The correlation remains above 0.8 up to the 5-hour time lag. This clearly shows that the evolution of regional production is rather slow. Moreover, simple regressive models, even Persistence, are likely to have very good performance when applied to regional forecasting, especially for horizons inferior to six hours. The correlogram also reveals a seasonal component of the time series. For lags 24, 48, 72 and 96 the autocorrelations presents peak values. This clearly indicates that there is a twenty-four hour cycle in the data. This is a well-known phenomenon in wind power and wind speed forecasting. Indeed, there is typically more wind power in the daytime than at nighttime.

With respect to the correlogram of the Irish regional production data (Figure 3.6), we can notice that the autocorrelations present a very similar pattern. The autocorrelation for the first time lags remains high (above 0.8), while they remain significant even for the higher time lags. Actually, the autocorrelations become insignificant at lag 150 for the Danish case and lag 800 for the Irish case. In the Irish case, the 24-hour seasonal cycle is also apparent.

When comparing the correlograms of both case studies, we can notice that the autocorrelation values for the first lags are higher for the Danish case. The fact that the Irish production data presents faster evolutions is possibly linked to the “size” of the case study. There are far fewer wind farms and much less installed capacity in the Irish case. This can explain that the Danish production presents higher “inertia” than the Irish production. However,

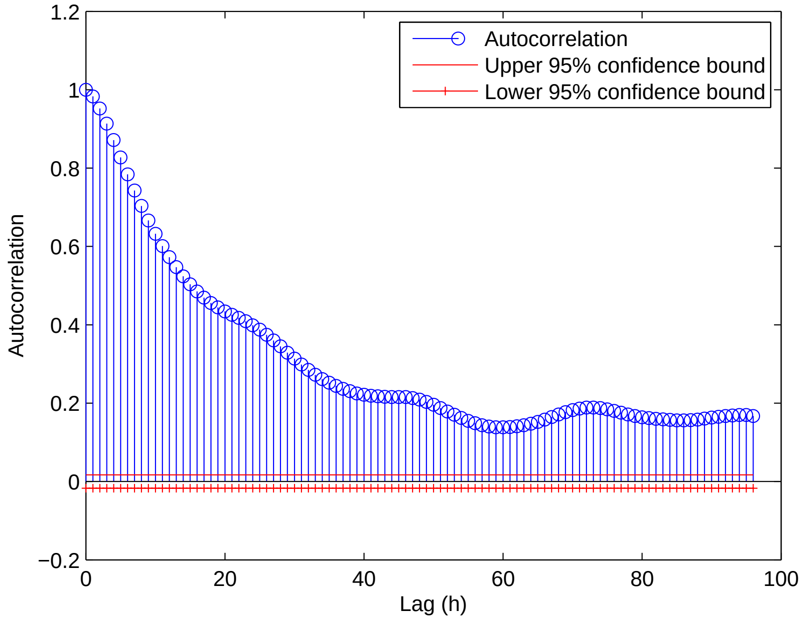


FIGURE 3.6: *Sample correlogram of the Irish production time series. The correlations are computed up to a lag of 96 hours.*

for higher time lags, starting at lag 24, the Irish regional production presents higher correlation values. This can be explained by the fact that the wind farms in the Irish case are geographically very concentrated. Indeed if wind farm 6 is not considered, the mean distance between wind farms in the Irish case is 92 kilometres and the maximum distance between wind farms is 249 kilometres, whereas in the Danish case, the mean distance between wind farms is 127 kilometres and the maximum distance is 296 kilometres. The Irish production is linked to the weather situation over a small area. On the contrary, the Danish production results from wind farms spread more evenly, hence presenting more diverse production patterns, which lead to lower correlation for higher time lags.

From the study of the moving averages and variances , as well as the correlograms, it is clear that the regional production is a non-stationary process. In order to deal with this property, implementing adaptive forecasting models seems necessary. Indeed, such models re-estimate their parameters in order to take into account evolutions in the predicted variable. Another conclusion, which can be drawn, is that Persistence is expected to have very good performance when predicting regional production.

3.6.2 Characterizing the statistical smoothing of regional production

In this section, we apply the experimental framework described in subsection 3.4.1 to the Danish and Irish case studies. The aim of this study is to examine the behaviour of the statistical smoothing of regional production, quantified by the smoothing factor S, when considering the aggregated production of wind farm combinations. The wind farm aggre-

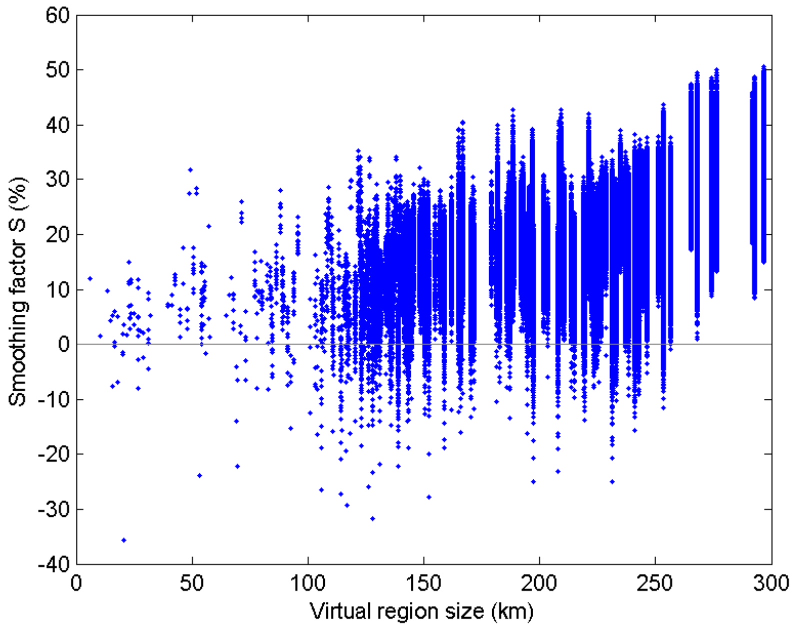


FIGURE 3.7: *Smoothing factor S versus virtual region size for the Danish case.*

gations are considered here as virtual regions. We examine the evolution of the smoothing factor with respect to the size of the virtual region under consideration, the relative geographical distribution of the wind farms in the region (quantified the size of the region and the mean distance between wind farms) and the number of wind farms in the region.

The results obtained being very similar for both case studies, only those of the Danish case (more than 8 million wind farm combinations) are presented here. The results for the Irish case can found in Appendix A.

In Figure 3.7 the smoothing factor S is plotted against the virtual region size, defined as the maximum distance between the wind farms in the aggregation. From this figure, it is apparent that the smoothing factor increases with the size of the region. For regions spanning less than 50 kilometres, the average smoothing is in the order of 5%. As the region size increases, the average smoothing increases toward values in the order of 30%. These results confirm those found in the literature and corroborate the theoretical results presented in subsection 3.4.1. However, a very interesting aspect is the existence of negative smoothing values. These are present for region sizes spanning a few tens of kilometres up to region sizes greater than 250 kilometres. Depending on which wind farms are considered in the virtual region, production smoothing may or may not occur, even for large regions. From this it can be concluded that although region size favours smoothing, the production variance of farms scattered over large regions will not necessarily lesser than the average variance of the individual wind farm productions. This in turn translates to regions where the upper bound on the forecasting error for the aggregated production is actually higher than that individual wind farms. From this we can reasonably suppose that cases can exist

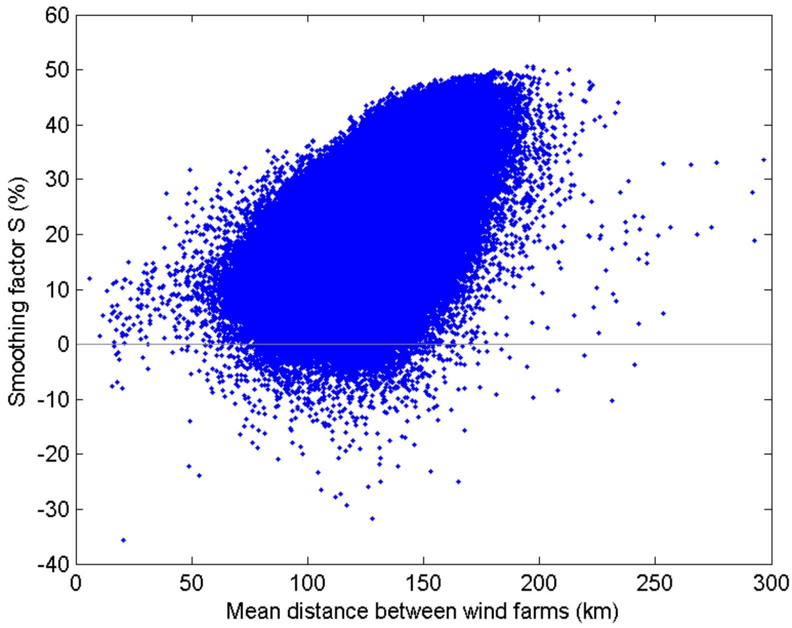


FIGURE 3.8: *Smoothing factor S versus the average distance between the wind farms in each combination.*

where forecasting the aggregate does not lead to lower normalized forecasting error.

Another factor that can be analyzed along with region size is the geographical dispersion of wind farms within the region. To approximately quantify wind farm dispersion we examine the evolution of the smoothing factor with respect to the average distance between the wind farms in the virtual regions. In Figure 3.8 the smoothing factor obtained for the wind farm combinations is plotted against the mean distance between the wind farms of the combination. From this scatter plot, we can conclude that the higher the average distance between wind farms is, the more likely smoothing will be positive. As with the region size, when the average distance increases, the number of cases where smoothing is negative decreases. In addition, even for regions with high mean distance, negative smoothing can also occur. These negative smoothing values can be attributed to an uneven distribution of wind farms within the region. However, the average distance between wind farms does not completely capture the relative dispersion of the wind farms in a region. A better evaluation can be obtained by comparing the mean distance between wind farms to the region size. The assumption made here is that if the mean distance is large when compared to the region the wind farms can be considered to be evenly distributed and unevenly distributed if the mean distance is small with respect to the region size.

In Figure 3.9 the smoothing factor is plotted against the mean wind farm distance and the region size. The points located on the diagonal correspond to the two-farm combinations. Indeed, for the two-farm combinations the average distance between wind farms is necessarily equal to the region size. With respect to the evolution of the smoothing factor, it appears that the higher smoothing values appear when the region size is large and the mean

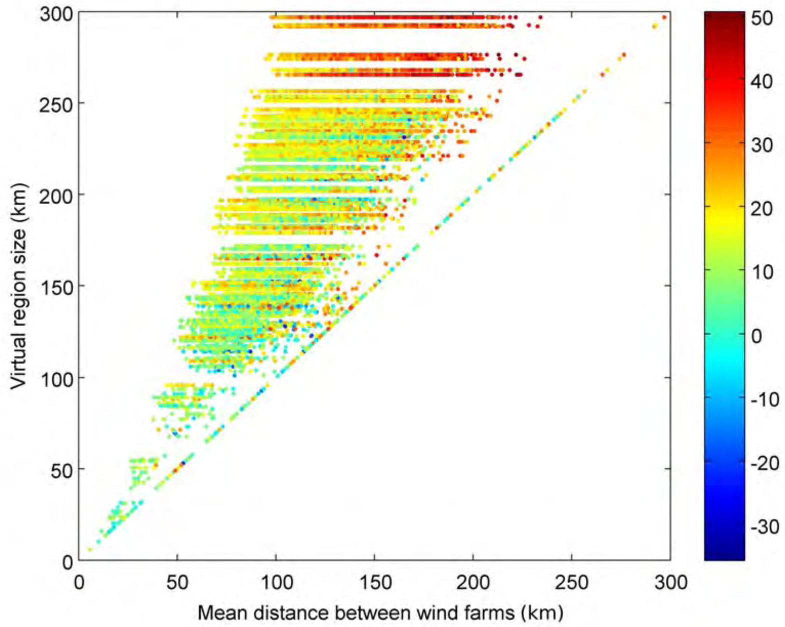


FIGURE 3.9: *Relation between region size and mean distance between wind farms on the smoothing factor. The colour bar on the right of the figure indicates the smoothing factor value for each wind farm combination.*

distance between wind farms is also large. Smoothing depends on both factors, however we can notice that region size seems to be more significant than the mean distance between wind farms.

From this analysis, it is clear that the limit derived in the hypothetical example considered in subsection 3.4.1 is valid with respect to the size of the region: when the size of the region increases the expected value of the smoothing factor increases. However, region size does not guaranty smoothing. For some large regions, negative smoothing can be observed. Furthermore, the geographical distribution of the wind farms within the region, as characterized by comparing the mean distance between wind farms to the region size, also plays an important role. If the distribution is uneven, i.e. if many wind farms are located close to each other then less smoothing will take place. Significant smoothing is observed when wind farms are distributed uniformly over a large area.

Concerning the influence of the number of wind farms on the smoothing factor, we plot the smoothing factor against the number of wind farms that comprise each combination (see Figure 3.10). From this figure, it is clear that the number of wind farms within a region does not play a predominant role on the observed smoothing. Indeed the smoothing values observed for 10-farm combinations are not higher than those observed for 5-farm combinations. However, we can notice that the highest and lowest smoothing values respectively decrease and increase as the number of wind farms increases. This can be explained in part by the fact that the number of wind farms used in this case study is limited. Hence, after a certain cardinality, the wind farm aggregations that lead to high or low smoothing are com-

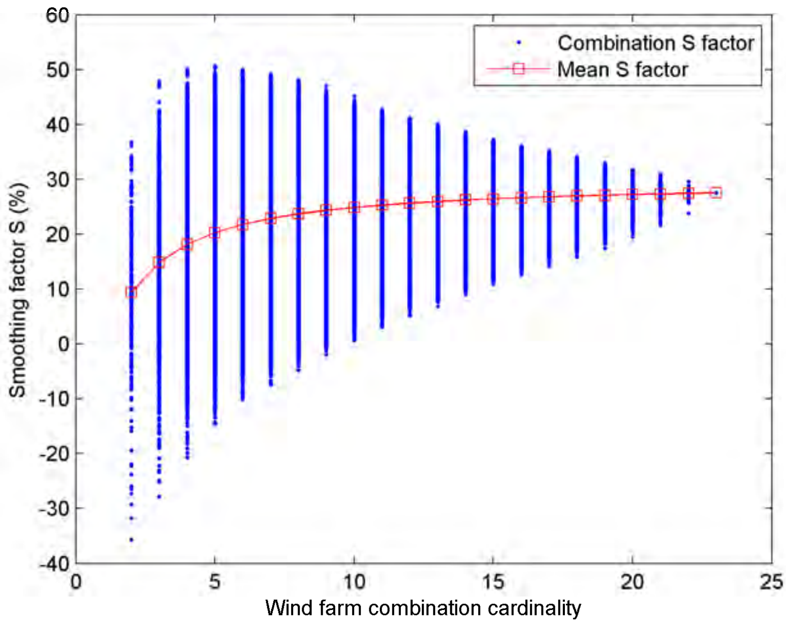


FIGURE 3.10: *Smoothing factor, and mean smoothing factor versus the number of wind farms in each combination.*

bined with other farms; this in turn mechanically leads to an aggregated production that presents either higher or lower aggregated production variance. Furthermore, we can observe that the mean smoothing factor for each cardinality increases as the number of wind farms increases and seems to attain a limit. This agrees with the theoretical limit found for the hypothetical case examined in subsection 3.4.1: the total production variance tends to a limit as the number of wind farms increases. The reduction of the spread of the observed smoothing values for large cardinalities can also be explained by the experimental framework. When the cardinality of the combinations exceeds half the number of available wind, farms the number of possible combinations for each cardinality decreases. Also the similarity between the wind farm combinations, in the sense of the number of wind farms that each combination shares, increases, making the resulting aggregated production variance and the mean single farm production variance similar for all combinations.

These results seem to suggest that the smoothing factor that can be observed over a region is constant. When considering a region with an important number of wind farms, spread rather uniformly, computing the normalized variance of the aggregated production of those farms will give a good estimate of the regional production which could be obtained if more farms where installed in the same region. To illustrate this we can compare the normalized aggregated production variance of the 23 Danish wind farms, and the normalized variance of the total Jutland-Funen production. For the 23 wind farms, the normalized variance equals 3.97% of the square nominal power and for the Jutland region, it equals 4.3%. The two values are quite close. From a practical standpoint, this constancy can possibly be used to estimate the production variance of a region from a relatively small number of single wind farm productions.

From the results presented in this section, the main conclusions that can be formulated are that the amount of smoothing effect is primarily dependent on the region size and the dispersion of the wind farms within the region. This agrees with the results published in [24] for the smoothing of regional forecast error. Furthermore, the number of wind farms considered does not play a predominant role. However, having many wind farms does seem to lead to an increase in the certainty of observing smoothing.

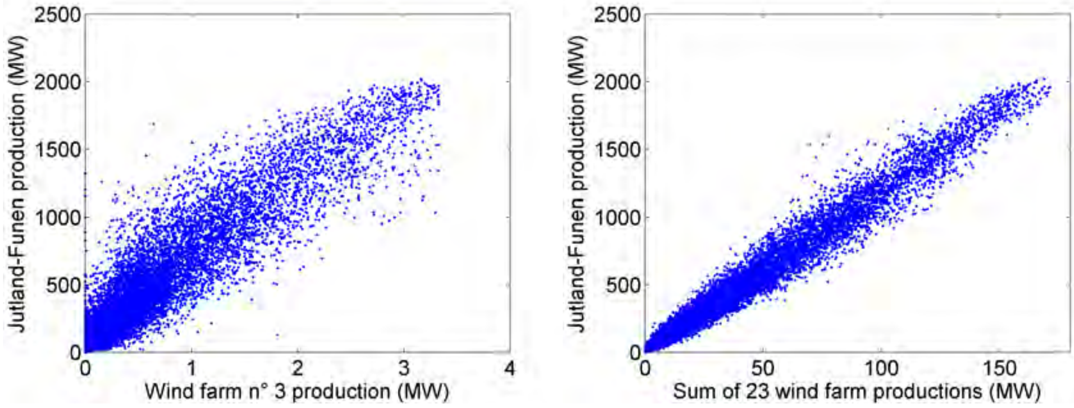
3.6.3 Characterization of the relation between single wind farm productions and regional productions

In this section, we will analyze the nature of the relation between the aggregated productions of a subset of wind farms and the regional production. The nature of this relation is of interest since on-line measurements of regional production are seldom available and regional forecasting models must therefore extrapolate the total production from a set of single wind farm production measurements. Knowing the nature of the relation between the production of a set of single wind farm productions and the total regional production permits to implement adequate extrapolation approaches. As described in subsection 3.4.2 one possibility is to examine the correlation between the aggregated production and the regional production. Before examining the correlation, we visually examine the regional production plotted against the aggregated production of subsets of wind farms. This allows us to determine the basic nature of the relation by simply eyeballing the scatter plot. We then investigate the strength of the linear relation between the aggregate production of different wind farm combinations and the aggregate production of all available wind farms. From this we then examine the dynamic nature of the relation using moving correlations in order to investigate the stability through time of the relation between the aggregated production of wind farm combinations and total production.

In Figure 3.11, two scatter plots are presented comparing the production of a single wind farm (in this case wind farm number 3) and of the aggregated production of the 23 Danish wind farms, to the total Jutland-Funen production[3]. The main aspect we can notice is that for both cases the relation to the total production possesses a very strong linear component. Furthermore, we notice that the linear relation seems to be much stronger for the 23 wind farm aggregate than for the single wind farm production.

The existence of a linear component in the relation between the single wind farm productions and the total production means that using correlation to investigate the strength of the relation is coherent since it measures the strength of the linear relation existing between two variables. In what follows, the correlation between the aggregated production of all the wind farm combinations and the aggregated production of all the wind farms available for each case study is examined. The wind farm combinations are defined as described in subsection 3.4.1. Of particular interest is the impact of the wind farm combination car-

[3]Note that for both case studies the general shape of this type of scatter plot is very similar, for this reason only the results for the Danish case are presented here.



(a) Single wind farm (WF3) (b) Sum of wind farms

FIGURE 3.11: *Aggregated production of a subset of wind farms versus total regional production for the Jutland-Funen case study. In figure (a) only the production of one wind farm is considered. In figure (b) the aggregated production of the 23 wind farms are considered.*

dinality on the correlation values. Indeed, in many cases on-line measurements of the total regional production are unavailable to TSOs. One of the questions which becomes crucial is how many wind farms are necessary to have a good estimate of the total production, and how can the wind farms be selected. The problem of on-line production estimation is also important in the frame of wind power forecasting. Indeed providing accurate forecasts, especially for small horizons (1 to 6 hours), requires having an accurate estimate of the past production.

In Figure 3.12 the results obtained for the Danish case are presented; again only the results for Danish case are presented since those for the Irish case are very similar. The results for the Irish case can be found in Appendix A. In this figure, the correlation values computed between the aggregated production of the wind farm combinations and the sum of the 23 wind farm productions are plotted against the wind farm combination cardinality. We can see that, as expected, when the number of wind farms increases the correlation increases, and reaches one when all the wind farms are included in the combination. The logarithmic shape of the correlation increase can be explained by the fact that as the cardinality increases the added information concerning the total production provided by an additional wind farm decreases. We can also notice that for the one-farm "combinations", most wind farms present high correlation to the total production, except one wind farm that seems to have an abnormally low correlation; this of course is case study specific. Interestingly, the two-farm combination correlations present a higher dispersion than the single wind farm correlations: the highest correlation value is higher and the lowest correlation is only slightly higher for the two-farm combinations than for the one-farm combinations. It is interesting to notice for the two-farm combinations, the highest correlation attained equals 0.96 (wind farms 13 and 17) and the lowest correlation equals 0.51 (wind farms 7 and 20). Hence, if the wind farms are adequately selected, even with few wind farms, it is possible to obtain high correlation values and hence good estimates of the total production. Further,

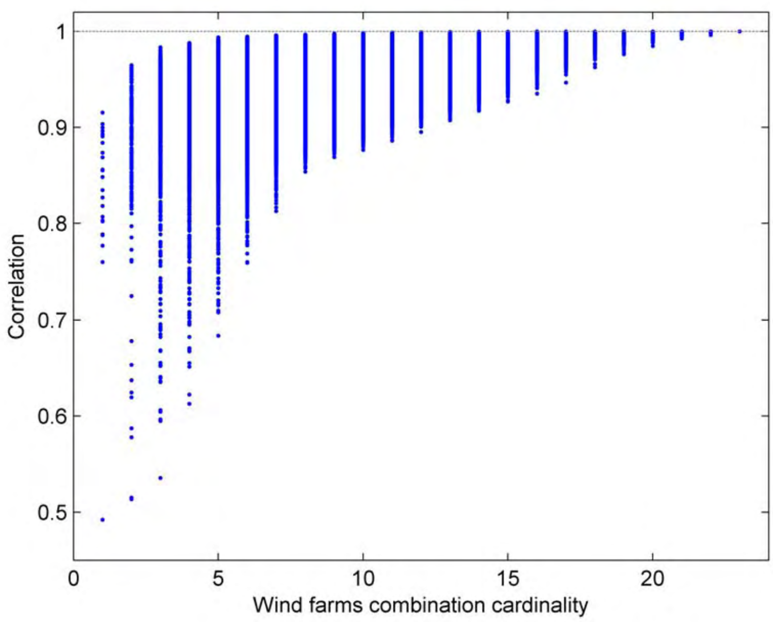


FIGURE 3.12: *Correlation between the aggregated production of wind farm combinations and the aggregated production of the 23 Danish wind farms. The correlation coefficients are plotted against the cardinality of the wind farm combinations.*

as the cardinality increases, the spread of possible correlation values decreases. With adequate selection, very few wind farms are necessary to provide a reasonable estimate of the production, for example the highest correlation obtained for a five-wind farm combination in both the Irish and the Danish case is superior to 0.99.

After examining the static relation of wind farm production to regional production, we analyze the nature of the relation from a dynamic standpoint. As mentioned in subsection 3.4.2, we use the moving correlation between the total regional production and the single wind farm productions. In Figure 3.13 the moving correlation between the Jutland-Funen production and the production of wind farm 3 and the aggregated production of the 23 wind farms is presented. The moving correlation is computed over 10 days, or 241 hours. As with the production time series, the correlations for both the single wind farm and the subset of wind farms are clearly non-stationary. Overall, the correlation variations are limited to a relatively small domain. Witnessing important variations, that is :high and null values for one wind farm, is rather rare, this only appears for a few wind farms in the two case studies available to us. Wind farm 6 in the Irish case presents several such correlation drops and even some negative correlation values. The values for this wind farm can be explained by the distance that separates this wind farms from all other wind farms. In most cases, however, these “extreme” events are comparable to the sharp drops present for wind farm number 3. These important correlation drops can often be explained by abnormal null production periods at the wind farms. These abnormally null productions can be caused either by wind turbine maintenance or by faulty production measurement systems. This

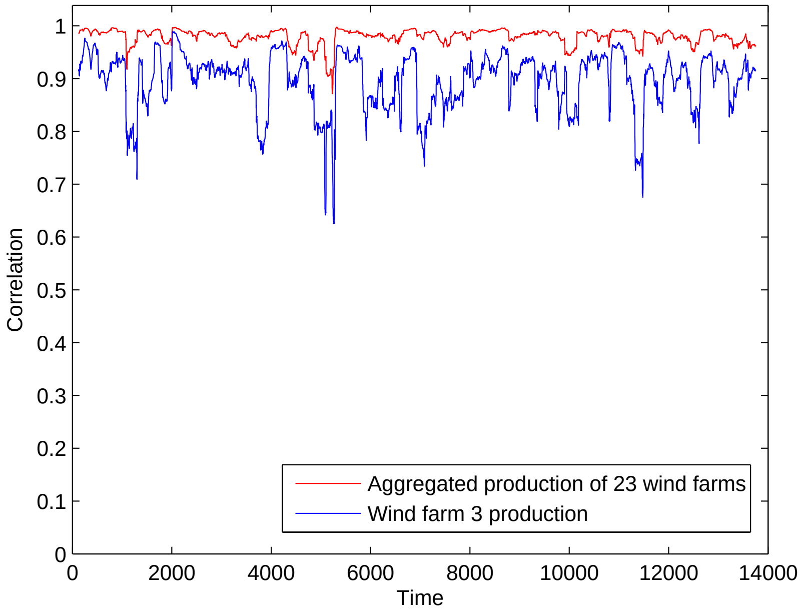
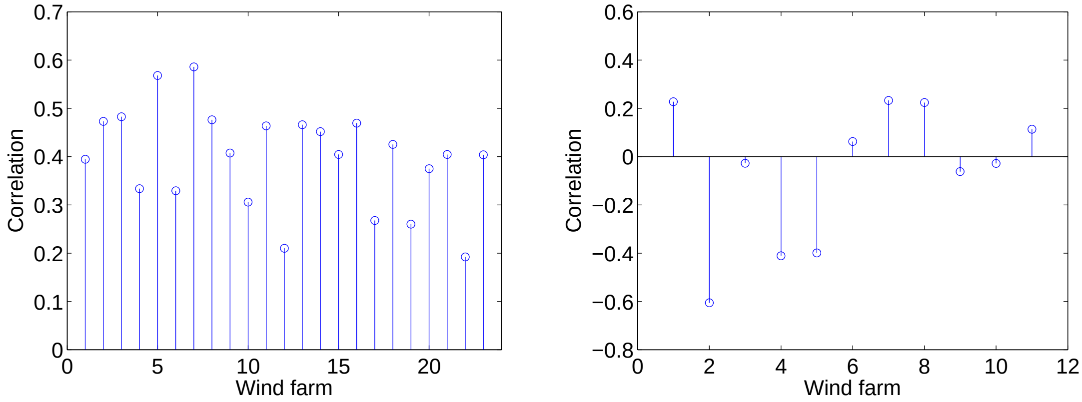


FIGURE 3.13: *Moving correlation between the Jutland-Funen production and wind farm 3 production, and between the Jutland-Funen production and the aggregated production of the 23 available wind farms. The moving average is computed over 241 hours (10 days).*

clearly illustrates the importance of possessing on-line wind turbine availability information. Also, in an operational setting detecting abnormally long constant production values might be necessary in order to detect malfunctions of the data acquisition systems.

When comparing the correlation values found for a single wind farm and the aggregated production of the 23 wind farms, we notice the same behaviour found in the static correlation study. As the number of wind farms considered in the aggregation increases, the moving correlation values increase, and the variations decrease. The reasons for this are obvious; if all the wind farms in the region are considered in the wind farm subset, the correlation will be maximal, and will not vary over time. In a way, by increasing the number of wind farms, the stationarity of the relation between the aggregate production and the regional production increases.

The non-stationarity of the relation between the wind farm productions and the regional production can further be examined by comparing the evolution of the moving correlation values to the moving averages of the wind farm productions. The moving averages computed for wind farm productions and regional productions show that all the time series are non-stationary. The moving correlations show that the strength of the linear relations existing between the wind farm productions and the total production are also non-stationary. Therefore, we investigate the possible existence of a relation between the wind farm production level and the correlation between the farms production and the total production. Here we compute the correlation coefficient between the moving average of the single wind farm productions and the moving correlation between the farm's production



(a) Danish case

(b) Irish case

FIGURE 3.14: *Correlation between the moving correlation between the single wind farm productions and the regional production, and the moving averages of the single wind farm productions. In sub-figure (a) the results for the Danish case are presented. In sub-figure (b) the results for Irish case are presented. In both cases the moving correlation and the moving averages are computed over 241 hours (10 days).*

and the regional production. Here we present results for both case studies in Figure 3.14. For the Danish case (sub-figure 3.14(a)) all the correlations are positive, and relatively high. This means that the correlation between the wind farm productions and the total production increases when the production of the wind farms increases. In other words when the wind speed is high over the region, the productions of the wind farms are highly correlated to the regional production. However, when the wind speeds are low, the productions of the wind farms are less correlated to the regional production. This could be explained by the influence of local effects on the wind farm productions. When the wind speeds are low, the correlation between wind farm productions decreases, because local effects have a high impact on wind farm production. However, when wind speeds are higher, local effects are less noticeable which translates into higher correlation between the farm productions. This phenomenon is evident in the Danish case, however, in Ireland (sub-figure 3.14(b)), this is not the case. In Ireland, the correlations are not all positive and the level of correlation varies greatly from farm to farm. This difference between case studies might be explained by the difference in terrain complexity. The farms in Denmark are located on flat terrain whereas farms in Ireland are located in much more complex terrain. However, factors such as faulty data, or the limited number of wind farms available might also play a significant role in this difference.

The conclusions that can be drawn from this section are primarily that for small regions, the relation between the single wind farm production and the regional production possesses a strong linear component. When the aggregated productions of wind-farm subsets of increasing cardinality are compared to the regional production, the strength of the linear relation increases. Another aspect that deserves some attention is that for regions with flat terrain, the evolution of the linear relation seems to be influenced to some degree by the level of production of the single wind farms.

3.6.4 Characterization of the relation between regional production and numerical weather forecasts

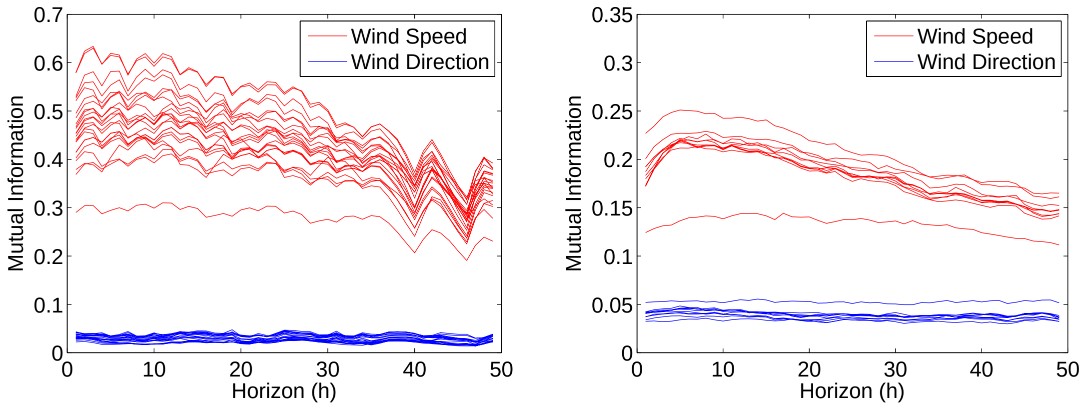
In this section, we investigate the strength of the statistical dependence between the explanatory numerical weather prediction variables and the total regional production. Because this relation is clearly non-linear, we use the mutual information to measure the dependence strength (see subsection 3.4.3). Because NWP forecasts are provided for several horizons into the future, we compute a per horizon mutual information. In this way, differences in the information provided by an NWP variable over the different forecast horizons can be detected.

Given the number of different types of NWP variables provided in both case studies (wind speed, wind direction, temperature, etc.) we limit the examination to the two types of variables which are provided in both case studies: wind speed and wind direction. These two types of data are the most commonly used in the wind forecasting community. In addition, the only altitude level common to both case studies is the 10 meters a.g.l. level. Since the aim of the present study is primarily to establish common traits between regional forecasting cases, we limit the examination to this level.

In Figure 3.15, the mutual information between the total regional production and both types of variables is plotted for the two case studies. In both cases we can notice that the information provided by the wind speed predictions is far superior to that provided by the wind direction forecasts. In the Danish case (figure 3.15(a)) the difference between both types of data is more pronounced than in the Irish case (figure 3.15(b)). This difference could be due to differences in the NWP model parameterization for both cases. Another explanation could be the difference in terrain complexity. Ireland has terrain that is more complex, this can possibly lead to a higher dependence of production levels on wind direction.

Another common trait in both cases is the shape of the mutual information curves for the wind speed forecasts. In both cases, the mutual information increases up to the 6 to 10 hour horizon, and then steadily decreases. This is more pronounced in the Irish case. This shape can be explained by the fact that NWP models are usually tuned to provide accurate forecasts for “medium term” horizons between 6 and 12 hours ahead. Hence, the mutual information increases for the first horizons and then naturally decreases because of the inherent randomness of atmospheric processes. It is interesting to notice that this reduction in mutual information is much less pronounced, if at all, for the wind direction forecasts. Because the information provided by the wind direction is very low, that is, the wind direction and the total production present little dependence, an increase in the prediction error of wind direction will not greatly influence the mutual information values found for these variables. The influence of the prediction horizon is very slight for the Irish case and barely discernible in the Danish case.

We can also point out the presence of regularly recurring “bumps” in the Danish wind



(a) Danish case

(b) Irish case

FIGURE 3.15: *Mutual information between regional production and NWP variables. Two types of NWP data are plotted: wind speed 10 m. a.g.l. and wind direction 10m a.g.l. In sub-figure (a) the results for the Danish case are presented. Each curve corresponds to the NWP variable provided for one of the 23 wind farms. In sub-figure (b) the results for the Irish case are presented. Each curve corresponds to the NWP variable provided for one of the 11 wind farms.*

speed mutual information values as well the two downward spikes present at the 39 and 45-hour horizons. The recurring oscillations span four horizons. The reason for their presence is probably that the NWP model used to compute the forecasts provides wind speed values with a three-hour time step; the hourly values provided between these values are interpolated in some way. This interpolation must be of lesser quality for horizons above 39 hours, which explains the important variations observed for these horizons.

In both case studies, some variables clearly have above or below average mutual information values. In the Danish case, this is particularly evident for the wind speed forecasts where the mutual information for one wind farm is clearly below those of the other wind farms. This curve represents the mutual information between the wind speed forecasts provided for wind farm 7 and the total production. In the Irish case, two wind speed mutual information curves stand out. One curve is clearly above all others and another is clearly below all other curves. The curve above all others corresponds to data provided for wind farm 4, and the curve below the others corresponds to the wind speed data provided for wind farm 6. In the Irish case we can also notice that one wind direction curve is well above the other wind direction curves. This curve corresponds to the wind direction data provided for wind farm 6. In both cases, and for both data types, the “odd” behaviour observed for these farms can be explained, to some extent, by the location of the wind farms. When examining the location of wind farm 6 in the Irish case (see Figure 3.1) and the location of wind farm 7 in the Danish case (see Figure 3.2), we can notice that both wind farms are in somewhat off-centre locations with respect to the other farms. This is particularly evident for wind farm 6 in the Irish case. This off-centre location can explain the fact that the meteorological data provided for these farms provides less information concerning the total production than information provided for wind farms situated in locations that are more central. This can explain the above average mutual information found for the wind speed provided for wind farm 4 in the Irish case. We can notice that wind farm 4 is located near wind farms: 2, 5, 7 and 8. Together, these five farms represent more than 64% of the installed capacity of the case study. Hence the wind speed provided for wind farm 4 likely provides much information concerning at least 64% of the installed capacity and it is likely that it also provides good information regarding the production at farm located farther south.

From the analysis of the information provided by the different NWP variables, it is clear that some types of variables seem to be potentially more useful than others. In this sense, in both case studies wind speed forecasts provide more information on the regional production than the wind direction forecasts. Also, we have seen that all NWP “sources”, that is wind farms for which NWP forecasts are available or, more generally, the NWP model grid points, do not provide the same amount of information regarding the regional production. Because of these differences, selecting the input variables to a regional forecasting model is clearly necessary in order to maximize the amount of information provided while limiting the amount of input variables taken into account. We can also mention that using “regional” NWP forecast has been suggested in the literature [84]. Here we conduct an evaluation on the feasibility of such an approach.

Regional NWPs as such are not usually available. Here we test the possibility of using the averages of the available NWP data as regional forecasts. Computing the mean wind speed is rather strait forward, however computing an average wind direction, is a bit trickier. To compute the average wind direction we propose to transform the polar wind direction θ available to us into Cartesian coordinates x and y.

$$x = r \cos\theta \tag{3.29}$$

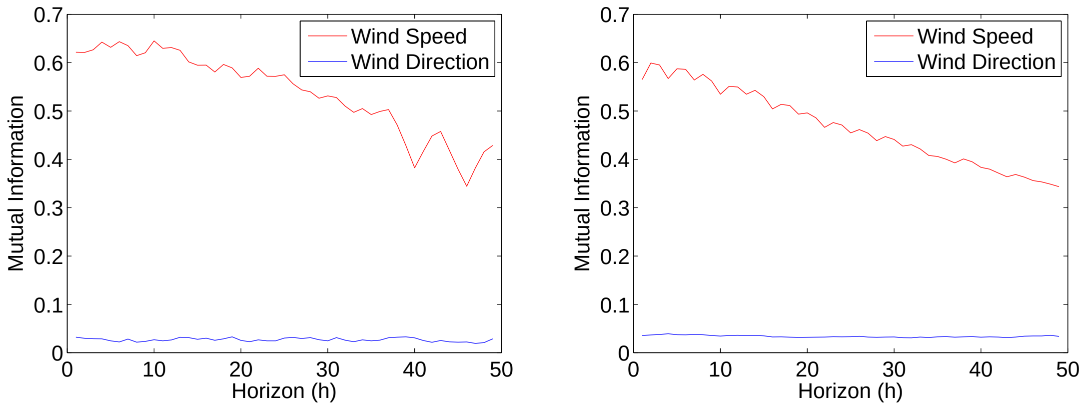
$$y = r \sin\theta \tag{3.30}$$

Here r is taken equal to 1. The average is computed on the Cartesian coordinates which is then converted into a polar direction θ according to:

$$\theta = \arctan \frac{y}{x} \tag{3.31}$$

Note that special care must be taken when x is null. In this case the direction equals $\pm 90^\circ$ depending on the sign of y.

In Figure 3.16 the horizon-dependent mutual information between the total regional production and the average wind speed, and average wind direction are presented for both case studies. We notice that for both cases the values for wind speed and wind direction are very similar: the wind speed values are slightly higher for the Danish case, and the wind direction values are slightly higher for the Irish case. In both cases, the average wind direction provides much more information than the wind direction. The shape of the average wind direction curves is very similar to the shape of the mutual information curves found for the single point wind speeds presented in Figure 3.15. We can also notice that the mutual information values found for the average data are similar to those found for the single point data for the Danish case. Interestingly, for the Irish case, the mutual information values found



(a) Danish case

(b) Irish case

FIGURE 3.16: *Mutual information between regional production and average NWP variables. Two types of NWP data are plotted: average wind speed 10 m. a.g.l. and average wind direction 10m a.g.l. In sub-figure (a) the results for the Danish case are presented. In sub-figure (b) the results for the Irish case are presented.*

with regard to the average wind speed are much higher than those found for the single point wind speed forecasts.

The results lead us to conclude that using “regional” NWP forecasts obtained through the averaging procedure described above can possibly be used to compute regional wind power forecasts. The information provided by the averaged NWPs are comparable to those provided by the individual NWP forecasts. It is highly likely that power forecasts derived from these averaged forecasts will have performance levels comparable to forecasts derived from purely single point NWP data.

3.7 Conclusion

Characterizing regional forecasting is not a simple matter. Indeed, many factors come into play and are often interdependent. In the work presented here we have looked at regional forecasting from a classic time series perspective, we have also investigated the regional smoothing effect. We then examined the relations existing between the regional production and the explanatory variables: single wind farm production measures and NWP forecasts available for certain farms. The evaluation was conducted on two case studies with the aim of determining common traits. However, it would be useful to apply the characterization framework presented here to other cases, such as Spain, to further establish general regional forecasting characteristics.

From a purely time series perspective it is clear that the regional production is a non-stationary process. This is a well know property of single wind farm production; however in the frame of regional production the non-stationarity is compounded by the fact that “structural” changes may occur such as the evolution of the installed capacity in a region. This can be problematic for forecasting models since the nature of the predicted time series evolves over time. A possible way of dealing with non-stationarity is to implement adaptive forecasting models. Such models periodically re-estimate their parameters in order to take into account evolutions in the relations between the explanatory and forecast variables. In addition, given the high autocorrelation values found for lags between 1 and 6 hours it is likely that Persistence will have very good performance, and it might prove difficult to provide forecasts that outperform it in the short-term horizons.

Characterizing the regional smoothing effect is not trivial. Here, we have examined the evolution of regional production variance as a function of different parameters. However, many parameters are inter-dependent. For example, the total installed capacity is clearly linked to the number of wind farms in a region and both are involved in the computation of the normalized regional production variance. Although difficult to establish, some conclusions can be formulated. The main one is that the amount of smoothing effect is primarily dependent on the region size and the dispersion of the wind farms within the region. The number of wind farms considered does not seem to play a predominant role. However, the presence of many wind farms does seem to lead to an increase in the likelihood of observing smoothing.

When examining the single wind farm production measurements as explanatory variables, our main focus was to determine to possibility of estimating the total production from partial measurements. As we have seen, on-line production measurements are mainly useful when computing short-term forecasts. Hence being capable of providing a good online estimate of the production is necessary to increase the accuracy of the first few forecast horizons. The conclusions that can be drawn from our results are that for small regions, the relation between the single wind farm production and regional production is highly linear. As the number of available single wind farm production measurements increases, the strength of the linear relation between the aggregated production and the total production increases. However, for a given number of available wind farms, some wind farm combinations can provide a better estimate of the current production. If the number of wind farm measurements that can be taken into account by a forecasting model is constrained, carefully selecting the production measures will be necessary. We also found that the relation between single wind farm production and regional production is non-stationary. This again argues for the possible necessity of implementing adaptive forecasting models. In this case, the models should be capable of dynamically selecting the production measurements on which to base their forecasts. To achieve this, one possibility would be to use combination models, like the one described in [65], to combine the forecasts provided by regional prediction models which use different input variables.

The other main type of input variables that we consider is NWP forecasts. From the analysis of the information provided by the different types of NWP variables, it is clear that some types are likely to be more useful than others are. Furthermore, as the prediction horizon increases, the strength of the statistical dependence between the NWP variables and the regional production decreases. This will translate into a forecast error increase for higher horizons. Also, we have seen that not all NWP model grid points provide the same amount of information regarding the regional production. We must bear in mind that statistical forecasting models are subject to parameter estimating difficulties when the number of input variables is too important. To avoid such problems, the number of input variables should be kept as small as possible. Because all NWP grid points do not provide the same amount of information on the regional production, selecting the input variables is clearly necessary in order to maximize the amount of information provided on the power production while limiting the number of input variables. One possibility that we investigated to keep input dimensionality low was to use "regionalized" NWP forecasts obtained by averaging the available single point NWP forecasts. We found that the information provided by the averaged NWPs is comparable to that provided by the individual NWP forecasts. Because of this, the quality of regional production predictions derived from these forecasts will be further investigated in the following chapters.

CHAPTER 4

Regional Forecasting Models

Abstract

The aim of this chapter is to examine regional forecasting models and provide guidelines toward their implementation. To this end, the regional wind power forecasting problem is defined from a statistical learning perspective. From this perspective existing regional wind power forecasting models are examined and three basic base-line model combinations schemes are identified. The work then concentrates on characterizing the performance of different base line models used in these three approaches. Two base-line models are used in this comparison: the state-of-the-art Fuzzy-Neural Network and a novel model named Regressive Power Curve (RPC) model. This model's performance is first validated by comparing its performance to that of other state-of-the-art models on three case studies of varying terrain complexity. The two base line models are then used to investigate the relative strengths of the different approaches and to examine the influence of using different base-line models in these approaches. This study is carried out on the Irish and Danish case studies described above.

4.1 Introduction

THERE exist several real-world applications where end-users, such as utilities or TSOs, operationally use regional wind power forecasts. The prediction systems are based on existing wind power forecasting models appropriately configured and combined to take into account the geographical nature of the problem and the dimensionality which stems from the availability of a wealth of explanatory variables.

These applications have been described to a certain extent in the literature but the available literature only covers specific cases. Most application function by using either physical or statistical models for single wind farm forecasting and ultimately combine their outputs through the use of statistical models. The models used in most cases result from research previously carried out to provide single wind farm forecasts. The outputs of these models, here after referred to as base models, are usually combined using a simple sum or, in cases where extrapolation is necessary, with a more advanced statistical model. However, in the literature, no guidelines or justification for the use of one model combination over another has been proposed. Also, no comparison of the performance obtained through the use of different base models is available.

The aim of the work presented here is to provide a more systematic view of the regional forecasting problem and ultimately to provide guidelines for the future implementation of regional forecasting applications. We concentrate on two main objectives. First, we determine from the literature basic base-model combination approaches and determine their relative merit on different case studies. We then concentrate on the impact of using different base models in identical combination approaches. Through this study, we will highlight some of the properties of regional forecasting, and provide model developers with guidelines with respect to the selection of base models and the selection of base model combination approaches.

The first part of the chapter introduces the regional forecasting problem from a statistical learning point of view and some notations are introduced to ease further discussion. The second part presents the basic forecasting approaches that can be identified from the literature and some examples from the literature are presented to illustrate the real-world use of these approaches. The justifications for the use of the combination methods are then discussed. We then introduce the base models that will be used in our study. One is the fuzzy-neural network (F-NN) based model developed by Kariniotakis [46], the other is a model proposed in this thesis which is based on piece-wise regression and specifically designed to have short computation time. This model is referred to as RPC model. In the fifth part, the different case studies that are examined are presented. In this part, a section is devoted to the validation of the RPC model. Indeed, it is necessary to ensure that this model possesses forecasting qualities comparable to those of other state-of-the-art forecasting models. The results of the study are then presented in detail in the last part of the chapter and general conclusions are formulated.

4.2 The Regional Forecasting Problem

Regional wind power forecasting may be required by end-users such as TSOs, wind farm operators, or electricity traders. Depending on the end-user the scope of the forecasts as well as the available data will be different. For example, an electricity trader will be interested in a forecast of the aggregated production of a few wind farms for which he possess SCADA measurements and NWP forecasts, whereas a TSO will be interested in a forecast of the aggregated production of all wind farms in a network region, in which case SCADA data and NWP forecasts may not be available for all wind farms. In the solutions open to the forecaster will be different. However, both cases can be seen as a general statistical learning problem. In this section, we will present the regional wind power forecasting problem as a statistical learning problem and we will introduce some notations that will allow further discussion in the next section.

In regional forecasting the focus is on the aggregated power P^{reg} of a set $\mathcal{A} = \{wf_1, ..., wf_M\}$ of M wind farms. The aggregated power at time t can be written as:

$$P^{reg}(t) = \sum_{i=1}^{M} P_i^{wf}(t), \tag{4.1}$$

where $P^{reg}(t)$ is the total regional power output at time t and $P_i^{wf}(t)$ is the power output of farm wf_i belonging to $\mathcal{A}$, at time t. In the following, the terms: region, group of wind farms, and set of wind farms are considered equivalent.

In regional forecasting the objective is to find a forecasting model f^{reg} such that:

$$\hat{P}^{reg}(t + k \,|\, t) = f^{reg}(X_t), \tag{4.2}$$

where: $X_t = \{x_{1,t}, \ldots, x_{M,t}, y_{1,t}, \ldots, y_{L,t}\}$ is the set of input data available at time t. Here $x_{i,t}$ denotes the input data, available at time t, which can be linked to wind farm wf_i. This data can include on-line production measurements, wind speed and direction measured at the wind farm and NWP forecasts provided for the site of the wind farm. Here $y_{j,t}$ denotes additional input data not specifically linked to a wind farm. This can include measurements of meteorological variables from measuring stations and NWP forecasts provided for locations other than those of wind farms in $\mathcal{A}$.

The aim is then to find f^{reg} given X_t that minimizes a loss function L. In the case of a least squares regression model, the aim is to find f^{reg} which minimizes:

$$L(f^{reg}, X_t) = \sum_{t=1}^{N-k} e(t + k \,|\, t)^2, \tag{4.3}$$

where N denotes the size of the data set over which the model performance is computed and $e(t+k \,|\, t)$ the model error defined in Equation 2.2. It must be noted that many loss functions can be defined depending on the forecasting problem. Often the least squares method is employed, however the aim might be to minimize the sum of absolute errors, the error bias, or the error entropy [103]. These loss functions are the ones most commonly used in the time series forecasting literature. This results from the fact that they are rather general loss functions, which aim at minimizing one or several moments of the error probability distribution function. Note that specific loss functions could be defined for the wind power forecasting problem in terms of avoided costs for the power system, or avoided penalties when considering an electricity market. However computing such loss function can prove to be a complicated problem in itself [80, 94, 104, 105] and would not necessarily reveal the strengths and weaknesses of forecasting models in a general setting. To remain general, we will only consider loss functions that relate to the performance estimation criteria defined in section 2.3.

One important aspect of regional forecasting is the availability of explanatory data. In order to ease the discussion on the available data we introduce the concept of reference wind farms. In addition, as mentioned in the last section, regional forecasting models often result from the more or less complex combination of base models. In order to allow the discussion of the structure of regional forecasting models we introduce notations for single wind farm models and sub-area models.

If we consider that for all wind farms in $\mathcal{A}$, $x_{i,t}$ consists of both on-line power measurements and NWP forecasts, then regional forecasting can be seen as a generalization of the single wind farm case. However, in practice, input data is seldom available for all wind farms. In this case, the forecasting problem can be said to lack information. Consequently the farms in $\mathcal{A}$ can be subdivided according to the characteristics of $x_{i,t}$. Let us define :

$$\mathcal{A}_v = \{wf_i \in \mathcal{A} \,|\, x_{i,t} = \emptyset\} \tag{4.4}$$

$$\mathcal{A}_r = \{wf_j \in \mathcal{A} \,|\, x_{j,t} \neq \emptyset\}, \tag{4.5}$$

where $\mathcal{A}_v$ is the subset of wind farms belonging to $\mathcal{A}$ for which no input data is available and $\mathcal{A}_r$ is the subset of wind farms belonging to $\mathcal{A}$ for which input data (NWP data and/or on-line measurements) is available. Wind farms belonging to $\mathcal{A}_r$ will be referred to as *reference wind farms*, and denoted as wf_{i_r}, $i_r = 1, \ldots, M_r$ and wind farms belonging to $\mathcal{A}_v$ will be denoted as wf_{i_v} with $i_v = 1, .., M_v$. In this case:

$$X_t^r = \{x_{i,t} \,|\, x_{i,t} \neq \emptyset\}, \tag{4.6}$$

is the set of available input data linked to the wind farms belonging to $\mathcal{A}_r$ for the regional case, and the problem is to find f^{reg} a forecasting model such that:

$$\hat{P}^{reg}(t + k \,|\, t) = f^{reg}(X_t^r, Y_t), \tag{4.7}$$

where $Y_t = \{y_{1,t}, \ldots, y_{L,t}\}$ corresponds to the input data not directly linked to the wind farms in $\mathcal{A}$. Note that in cases where $\mathcal{A}_v \neq \emptyset$, it is highly likely that $Y_t = \emptyset$. In such cases, the forecasting problem lacks information, therefore, it cannot be considered as a direct generalization of the single wind farm problem. To provide good forecasts the forecasting model must possess extrapolation capabilities in order to exploit to the fullest the limited amount of information available. Models designed to deal with such cases will hereafter be referred to as *upscaling* models.

As we have mentioned, many different model configurations are possible. To describe these configurations we introduce additional notations. To distinguish single wind farm models and regional models, single wind farm models will be noted as f_i^{wf}. Sometimes a region can be subdivided into sub-areas. In this case, power forecasts for the sub-area are the result of a regional forecasting model. However, the sub-area power forecasts are usually used as input data to the regional power model. To avoid confusion, these sub-area models will be noted: f^{ar}.

To illustrate these two notations we can consider two trivial cases. Let us first consider a case where input data is available for all wind farms, and no additional data is available, i.e.

$Y_t = \emptyset$. Then, the regional forecast could be given by:

$$\hat{P}^{reg}(t+k\,|\,t) = f^{reg}(f_1^{wf}(x_{1,t}),\ldots,f_M^{wf}(x_{M,t})) = \sum_{i=1}^{M} \hat{P}_i^{wf}(t+k\,|\,t) \tag{4.8}$$

In this example the regional model simply computes the sum of the forecasts provided by the single wind farm forecasting models f_i^{wf}.

In the second example, let us consider the case where the farms in $\mathcal{A}$ are divided into d sub-sets or sub-areas $\mathcal{B}_j$ such that:

$$\mathcal{A} = \bigcup_{j=1}^{d} \mathcal{B}_j \quad \text{and} \quad \bigcap_{j=1}^{d} \mathcal{B}_j = \emptyset \tag{4.9}$$

the regional forecast could be given by:

$$\hat{P}^{reg}(t+k\,|\,t) = f^{reg}(f_1^{ar}(X_{1,t}^r),\ldots,f_d^{ar}(X_{M,t}^r)) = \sum_{j=1}^{d} \hat{P}_j^{ar}(t+k\,|\,t), \tag{4.10}$$

where $X_{j,t}^r$ is the available input data for the farms in sub-area j, and $\hat{P}_j^{ar}(t+k\,|\,t) = f_j^{ar}(X_{j,t}^r)$ is the forecast for sub-area j. As in Equation 4.8, the regional forecasting model f^{reg} is only the sum of the sub-regional forecasts.

4.3 Regional Modelling Approaches

In the frame of regional forecasting, as in the case of single wind farm production forecasting, it can be useful to distinguish different types of models. As with single wind farm models the distinction could be the theoretical basis of the base model being used: physical or statistical. Indeed, in a case where input data is available for all wind farms in $\mathcal{A}$, a possible forecasting solution could be to use the trivial regional model described in Equation 4.8, where individual forecasts are computed for each wind farm. The regional forecasting model could then the sum of the individual forecasts. In this case the single wind farm models f^{wf} can be either physical or statistical models. By analogy, depending on the type of single wind farm forecasting model used, the regional model could be said to be either physical or statistical. When input data is available for all wind farms, a mix of statistical and physical base models could even be used. This would then lead to “hybrid” regional models.

However, most physical forecasting models use some statistical correction and can therefore be considered hybrid. Also, in most cases data is not available for all wind farms. In such cases, using the approach described above would not possible and some sort of statistical extrapolation would be necessary to compensate for the lack of information concerning the farms in $\mathcal{A}_v$. Here “statistical” can stand for a simple scaling factor applied to the sum of single wind farm forecasts or a more advanced model such as an artificial neural network. In such cases, the regional forecasting model can no longer be defined as being physical or statistical. Therefore, the physical versus statistical typology applied to single wind farm models cannot be directly transcribed to regional forecasting models in general.

A possible way of classifying regional forecasting models can be through the analysis of their "architecture". By architecture, we mean the way in which the model can be described as the combination of base models. Here, we use the notations defined in section 4.2. Indeed, a regional model f^{reg} can often be defined as the combination of sub-regional models f^{ar} and/or single wind farm models f^{wf}. By approaching regional forecasting models in this way, three *basic* configurations or approaches can be established:

- The *direct* approach. In this case, the regional model takes as input all available input data and directly computes the regional power production forecasts. This can be written as $f^{reg}(X_t^r, Y_t)$.
- The *cascaded* approach. In this case, the regional model uses as input single wind farm forecasts provided by single wind farm forecasting models. This can be written as $f^{reg}(f_1^{wf}(x_{1,t}), \ldots, f_M^{wf}(x_{M,t}))$. This approach is typically used by regional models that rely on physical single wind farm forecasting models.
- The *cluster* approach. In this case, the regional model uses as input forecasts provided by sub-regional forecasting models. This can be written as $f^{reg}(f_1^{ar}(X_{1,t}^r), \ldots,$ $f_d^{wf}(X_{M,t}^r))$. The sub-regional models can be either direct, cascaded or cluster models.

These three configurations are quite generic. They can be used to describe many existing regional models. However, these configurations are not the only ones that can be imagined, especially when considering combinations of statistical base models.

Here we use these three basic configurations to describe two state-of-the-art regional forecasting models. The first model is the upscaling algorithm included in the Previento forecasting system developed at Oldenburg university [57]. This system is based on a physical single wind farm forecasting model. The approach used in this upscaling model is a *cluster* approach where the sub-area models use a *cascaded* approach. The authors suggest dividing the considered region into a number of sub-areas. In each sub-area, a reference wind farm is selected and the physical model is used to compute the predictions of the power output for the farm. For each sub-area, a *cascaded* model then upscales the single wind farm forecasts to compute the sub-area power forecasts. The final regional power prediction is calculated by a *cluster* model from the sub-area power forecasts.

The second model is the model developed by DTU and described in [84]. A graphic representation of this model's configuration can be seen in Figure 4.1. This model uses a two-branch configuration. In the left branch, sub-area models noted $PP_{i,1}^{ar}$ use a *cascaded* approach to derive sub-regional forecasts from single wind farm forecasts computed using models designated by $PP_{i,j}^{wf}$. The sub-area forecasts are then combined by a *cluster* model, which in this case is a simple sum, to obtain the regional forecast noted as $\hat{P}_1^{to}$. The right branch of the model uses a *cluster* approach based on sub-area models $PP_{i,2}^{ar}$, which use

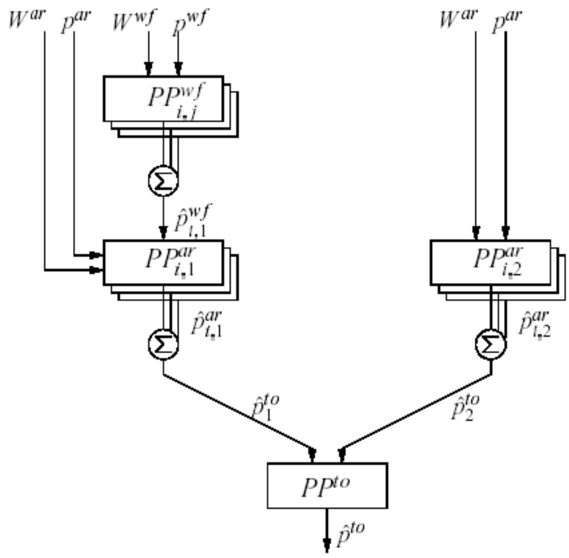


FIGURE 4.1: *Graphic representation of a regional forecasting model developed at DTU. The figure is taken from [84].*

the *direct* regional forecasting approach. The two regional power forecasts $\hat{P}_1^{to}$ and $\hat{P}_2^{to}$ are used as inputs to PP^{to}, which computes a weighted average of the forecasts from the two branches.

The upscaling model of the Previento system is a straightforward combination of two basic regional forecasting approaches. It illustrates the usefulness of the *cascaded* approach when physical single wind farm models are used as the basis for a regional prediction model. The example developed at DTU illustrates the complexity that can be attained by statistical models. This model uses all the basic configurations described above. Furthermore, this example clearly shows the nesting and combination possibilities of the different configurations. In the left branch a *cascaded* approach is used in combination with a *cluster* approach and in the right branch a *direct* approach is used in combination with a *cluster* approach. We must also note that the final combination model can be considered as a *cluster* model in which both sub-area *cluster* models represent the same area.

Clearly, an almost infinite number of configurations can be defined. Mixing physical and statistical models, as well as incorporating data from different sources at different stages of the forecast chain is possible. The use of a given configuration must however be justified. In the available literature, such justifications are mainly based on forecast accuracy considerations. For the Previento configuration described above [57], the authors aim at minimizing the forecast-error smoothing, which they define as $\sigma_{ensemble}/\sigma_{single}$, the ratio between the standard deviation of the regional forecast error ($\sigma_{ensemble}$) and the average standard deviation of a single wind farm forecast error (σ_{single}). Following the error smoothing study in [24], they argue that for Germany, the level of error smoothing does not decrease significantly when more than 50 reference wind farms are considered. Hence, in [57], the authors divide Germany into 45 sub-areas. For each sub-area, a reference farm is selected, then their upscaling algorithm is applied.

With respect to the model developed at DTU [84], the authors argue that the left branch of the model takes advantage of the auto-correlation present in the power production for horizons inferior to 12 hours, whereas the right branch takes advantage of the smoother nature of the sub-regional production and the fact that numerical weather models more reliably predict weather "patterns" over a region than weather conditions at a particular wind farm site. In essence, this means that the left branch, because it uses power measurements from individual wind farms as input, can more accurately predict power production for horizons below 12 hours. Conversely, the right branch considers aggregated power measurements, which necessarily provide less detailed information than individual farm measurements. In the same way, the NWP data used by the direct sub-regional models are averaged values upon the area and are consequently less "detailed" than the NWP data used in the left branch. This allows the right branch to more accurately capture the relation between future weather patterns and the smoother sub-area production. The two branches can be seen as exploiting both "high resolution" details (left branch) and "low resolution" trends (right branch) provided by the input data.

The two architectures described in this section as example, clearly show some of the issues that must be considered when building a regional forecasting model. The architecture definition method described in [57] for the Previento system concentrates on finding the amount of input data necessary to minimize the forecasting error. The architecture described in [84] for the second example concentrates on fully exploiting the information contained in the available input data. From this it is clear that the model structure is not independent of the expected accuracy. However, the model architecture can also be conditioned by more "mundane" factors such as end-user requirements or data availability restrictions. Indeed, end-users sometimes request particular architecture; this can be a cascaded approach for sub-regional models, the region being divided according to predefined market zones. In other cases, sub-regional production data might not be available. In that case using a cluster approach is not possible since no training data is available to tune the sub-area models. These factors can constrain the architectures a modeller can choose from. In the literature, no indication on the influence of the model architecture on forecast accuracy is available. Hence, examining the impact that model architecture will have on accuracy is necessary so that the best architecture can be chosen within the end-user's specific constraints. This aspect will be examined in the following section of this chapter. Input selection in the frame of regional forecasting will be examined in chapter 5.

4.4 Statistical Learning Models for Regional Forecasting

We have mentioned that regional forecasting models can be based on physical or statistical modelling approaches. In this section, we concentrate on the statistical regression models that constitute the core of statistical regional prediction models.

As mentioned above statistical models refers to models whose design aims at finding the dependence between variables, given a finite number of observations. The aim of such models is to "learn" this dependence. Here, the general learning problem is presented as described by Vapnik in [106]. It can be described through three components:

(i) A generator (G) of random vectors $x \in R^n$, drawn independently from a fixed but unknown probability distribution $F(x)$.

(ii) A supervisor (S) who returns an output value y to every input vector x, according to a conditional distribution function $F(y \,|\, x)$, also fixed but unknown.

(iii) A learning machine (LM) capable of implementing a set of functions $f(x, \alpha)$, $\alpha \in \Lambda$, where Λ is a set of parameters.

The problem of learning is that of choosing from the given set of functions $f(x, \alpha)$ the one that best approximates the supervisor's response. The selection of the desired function is based on a *training set* of l independent and identically distributed observations drawn according to $F(x, y) = F(x)F(y \,|\, x)$:

$$(x_1, y_1), \ldots, (x_l, y_l) \tag{4.11}$$

In the frame of wind power forecasting, the generator G can basically be considered to be the atmosphere which generates "random" weather conditions described by data vector x. The supervisor is then a wind turbine, a wind farm, or the wind farms in a region, which generate a power output y according to a relatively "unknown" distribution function. The learning machine is here the wind power forecasting model, whose task is to model the power output y from vector x. In essence, wind power forecasting models are designed to estimate the conditional distribution function $F(y \,|\, x)$.

To solve the learning problem, Vapnik describes the risk minimization problem, where the goal is to approximate the supervisor's response by selecting the function which minimizes the *loss*, or error, $L(y, f(x, \alpha))$ between the supervisor's response y to input x and the response $f(x, \alpha)$ computed by the learning machine. To minimize the expected value of the loss, the following *risk functional* is defined:

$$R(\alpha) = \int L(y, f(x, \alpha)) dF(x, y). \tag{4.12}$$

The goal is then to find $f(x; \alpha_0)$ that minimizes $R(\alpha)$ over the class of functions $f(x, \alpha)$, $\alpha \in \Lambda$, where $F(x, y)$ is unknown and the only available information is contained in the empirical training set defined in 4.11. This risk minimization problem is a generalization of the regional forecasting problem as described in section 4.2.

Much research has been conducted on the statistical learning problem. Rosenblatt developed the first learning machine, the Perceptron, in the early 1960s [107]. This was the first model capable of learning from experimental data, and was at first designed to solve pattern recognition problems. This experimental approach to statistical learning spurred research in the field. At the same time Vapnik and Chervonenkis started developing a "philosophy of statistical learning theory". The theoretical work they developed led to a new inductive principle called structural risk minimization in 1974, as opposed to the empirical risk minimization principle. In a nutshell, empirical risk minimization is intended for large sample sizes whereas structural risk minimization is specifically designed to deal with small sample sizes. Vapnik and Chervonenkis completed their work on the structural risk minimization principle by the early eighties [108]. Finally, in the late eighties, the final theoretical consistency conditions for the empirical risk minimization were found [109]. Based on this theory Vapnik and his team went on to develop a new type of learning model: support vector machines [110]. While Vapnik and others were developing the theory behind statistical learning, other researchers where focusing on the experimental development of learning machines. From this perspective, a gradient method named error back-propagation was proposed for training the hidden layers of neural networks. This breakthrough in learning from data and neural networks was independently discovered by several authors [111, 112]. More recently, learning models based on the theory of fuzzy sets proposed by Zadeh [113], have been developed. Although based on a very different paradigm, that is including human expert knowledge in a model, some types of fuzzy-logic-based models have been found to be equivalent to certain neural network models [114]. These learning models can however allow extracting knowledge from the modelled process, whereas interpreting neural network weights can be quite complicated.

The two major requirements a learning machine must fulfil are, as described in [106]:

(i) To estimate the desired function from a wide set of functions.

(ii) To estimate the desired function on the basis of a limited number of examples.

The first requirement relates to the modelling power of the learning machine, that is the type of function that can be modelled. From this point of view, many neural networks and other learning machines have proven to be, under some conditions, universal approximators. In other words, these models can model any function belonging to a wide set of functions. The second requirement relates to the ability of a model to "generalize". Generalization is the ability to correctly model the process of interest from the available learning set. Since learning sets are noisy, the learning machine must "filter" the noise to capture the underlying process. If the learning machine is not well designed, this can lead to underfitting or overfitting. In the case of underfitting, the model has "filtered" not only the noise in the data but also part of the underlying process. When overfitting occurs, the learning machine has incorporated some of the noise present in the learning data into the model of the underlying process. If the learning machine does not generalize well, its modelling performance will not be good.

The statistical learning theory initially developed by Vapnik aimed at providing theoretical tools capable of helping modellers design models that fulfilled both requirements stated above. The theory-based model Vapnik and his colleagues proposed were support vector machines (SVM). In theory, these models should be better than other, experimental, or empirical, models such as neural networks. However in [106], Vapnik discusses the differences in performance of different models on the digit recognition problem for an official benchmark case defined by the U.S. National Institute of Standards and Technology. When SVMs where first compared to the existing state-of-the-art neural network, they outperformed it. However, the neural network models were refined and this led to an equal error rate for both polynomial support vector machines and for neural networks. Vapnik then mentions another improvement on the neural network model that lead to even lower error by injecting a priori knowledge into the network. When the same thing was done with the SVM models, both modelling approaches once again had roughly the same performance. From this it is clear that both approaches, the theoretical and empirical, are capable of solving the learning problem, and that both methods benefit from the modeller's insight on the problem itself. Hence, although the objective merits of the learning machine are important when trying to design a learning machine for a given problem, the modeller's knowledge of the process to be modelled, as well as his knowledge of the learning machine he is employing, counts for a great deal in the results which can be expected.

In this chapter we aim at comparing two different learning machines, and their results, when used in different regional forecasting approaches. The learning machines we will use are the state-of-the-art fuzzy-neural network wind forecasting model developed by Kariniotakis [46]. The second model is one is proposed here and is based on a more "physical" modelling of the problem.

4.4.1 Fuzzy-Neural Network Model

The fuzzy-neural network (F-NN) model developed by Kariniotakis [46, 64] at the Centre for Energy and Processes is a Takagi-Sugeno fuzzy inference system. It derives its name from the fuzzy logic and neural network background that inspired its creation. In this section we present a summary of the corresponding part in [46] where a detailed description of the model, the parameter identification, as well as the architecture optimization are found. This will allow us to introduce these two latter concepts and compare the F-NN model to the model we propose.

This artificial intelligence (A.I.) model is data independent: it has been shown to outperform existing time series models on classic benchmark cases [46], and it can combine time series and forecast data as input data or use only forecast data with good accuracy.

This model, like many artificial intelligence models, is a black box model. No knowledge of the relations existing between the explanatory and dependent variables is necessary to implement it. Indeed, the aim of artificial intelligence techniques is to mimic the learning processes of the brain to discover the relations between the variables of a system. Many A.I. models are true black box models since very little can be learnt by examining the values of their parameters. However, as we shall see in the next section, the F-NN is an expert system. The model's learning phase basically builds a set of fuzzy rules describing the relations between explanatory and dependent variables. It has been claimed that by examining the rule-base, knowledge about the nature of the variable's relations can be gained. This however can prove difficult in the wind power forecasting problem where tens of rules are necessary in order to attain good forecasting accuracy.

Also, like many neural-networks, the F-NN uses supervised learning techniques [46].

This means that the tuning of the model's fuzzy rules is based on the difference between the expected output and the actual output of the model. These learning rules can be extended, with very few modifications, to become online adaptation rules. Because of this, if the expected output (in our case the measure of the regional wind power production) is available on-line, the model can adapt to changes such as the evolution of regional installed capacity, turbine aging, seasonal weather and environmental changes, and evolutions in NWP accuracy.

Defining the fuzzy model

The fuzzy model can be considered as an inference engine or expert system that uses a rule base to determine the desired output from a given input. The fuzzy rules can be expressed as:

$$\text{IF } \underline{x} \text{ is } A \text{ THEN } y \text{ is } B$$

where $\underline{x}$, y are linguistic variables and A, B are fuzzy sets. In the model developed by Kariniotakis the rules have the form:

$$\text{R : IF } x_1 \text{ is } A_1 \text{ and}, \ldots, \text{ and } x_n \text{ is } A_n \text{ THEN } y = g(x_1, \ldots, x_n)$$

where:

- $x_1, \ldots, x_n$ are real-valued variables representing input variables of the system defined in the universes of discourse $X_1, \ldots, X_n$ respectively.
- $A_1, \ldots, A_n$ are fuzzy sets.
- y is the variable of the consequence whose value is inferred. In the problem at hand it represents future wind power .
- $g(\cdot)$ is a function that implies the value of y when $x_1, \ldots, x_n$ satisfy the premise.

The function $g(\cdot)$ in the consequent part of the rules may be linear, non-linear or constant.

The interpretation of the fuzzy rules described above requires the evaluation of the premise of the rule as well as the evaluation of its consequence. To interpret the premise of the rule the sentence connective "and" must be interpreted. For the fuzzy sets $A_1, \ldots,$ A_n present in the premise of a rule and defined in the universes of discourse $X_1, \ldots, X_n$, the membership function $\psi_{A_i}(x_i)$ represents the degree to which the input value x_i corresponds to the linguistic term associated with fuzzy set A_i. In order to fulfil the premise, the inputs $x_1, \ldots, x_n$ must fulfil the linguistic terms associated with $A_1, \ldots, A_n$, hence the values of $\psi_{A_1}, \ldots, \psi_{A_n}$ must be conjunctively combined. This conjunction is achieved by defining a Cartesian product of $A_1, \ldots, A_n$ in $X_1, \ldots, X_n$. The Cartesian product is a fuzzy set $A = A_1 \times \ldots \times A_n$ in the product space $X_1 \times \ldots \times X_n$ with membership function $\psi_A(\underline{x}) = \psi_{A_1 \times \ldots \times A_n}(x_1, \ldots, x_n)$. To define the Cartesian product membership function any operator associated with a T-norm can be applied. In the present case the algebraic product operator is used:

$$\psi_A(\underline{x}) = \psi_{A_1 \times \ldots \times A_n}(x_1, \ldots, x_n) = \psi_{A_1}(x_1) \cdot \psi_{A_2}(x_2) \cdot \ldots \cdot \psi_{A_n}(x_n) \tag{4.13}$$

After evaluating the premise of a fuzzy IF-THEN rule, the fuzzy implication function $R : A \rightarrow B$, must be interpreted to determine the conclusion of the rule. Many implication functions have been defined in the literature. In the present model the fuzzy conjunction is used. The fuzzy conjunction is defined as:

$$\psi_R(x, y) = \psi_{A \rightarrow B}(x, y) = \psi_A(x) \tilde{*} \psi_B(y) \tag{4.14}$$

where $\tilde{*}$ represents a T-norm operator. In the present case the fuzzy model is a Takagi-Sugeno inference engine, hence the conclusion fuzzy set $\psi_B(y)$ is replaced by the $g(\cdot)$ function. The T-norm operator used to compute the fuzzy implication is the algebraic product. Hence in the present case the fuzzy implication function R can be written:

$$\psi_R(\underline{x}, y) = \psi_{A_1 \times \ldots \times A_n}(x_1, \ldots, x_n) \tilde{*} \psi_B(y) = \prod_{j=1}^{n} \psi_{A_j}(x_j) g(\underline{x}) \tag{4.15}$$

Note that using the algebraic product operator leads to Larsen's implication operator. Further, using this operator leads to a simple analytical expression of the rule's result. In order to simplify the following notations, the value of $\psi_{A_1 \times \ldots \times A_n}(\underline{x})$ will be noted w. Here w can be considered, from a connectionist or neural network point of view, as the weight or firing strength of rule R; it is the degree to which the premise of the rule is fulfilled.

We have seen how the output of a fuzzy rule is inferred in the fuzzy-neural model. However we have seen that the model is a fuzzy inference engine. Hence the final crisp output must be inferred from the outputs of the fuzzy rules defined in the rule base. In order to carry out the inference different methods are available. In the present model the following general procedure is employed:

1. The output fuzzy sets $B^1, \ldots, B^m$ of each rule are computed.

2. The output B^i of each rule is defuzzified in order to produce a crisp output.

3. The total output of the model is the weighted average of the crisp outputs of the rules.

In fuzzy model cases the centre value defuzzifier is generally used in step 2. Then the combination of the three steps, evaluation, defuzzification and weighing is termed the centre average deffuzifier. However, in the present model the rule output is already crisp as can be seen from Equation 4.15. Therefore, the only necessary steps are the weighing and averaging ones. Hence, the crisp output of the fuzzy model is given by:

$$f(\underline{x}) = \frac{\sum_{i=1}^{m} w^i \, g^i(\underline{x})}{\sum_{i=1}^{m} w^i} \tag{4.16}$$

where $w^i = \prod_{j=1}^{n} \psi_{A_j^i}(x_j)$ is the weight of rule R^i, $g^i(\cdot)$ is the consequent function of rule R^i, and m is the number of rules in the rule base.

From the general expression of the model given in Equation 4.16, we can conclude that this model is capable of *locally modelling* the relation between the input variables $\underline{X}$ and the expected output Y. In essence, the premises of the fuzzy rules can partition the input space defined by the variables in $\underline{X}$ into m "fuzzy" regions. When an input vector $\underline{x}$ is located in region k, then the firing strength of rule R^k will be high. This means that $\underline{x}$ strongly belongs to the fuzzy set defined by the Cartesian product of $A_1^k, \ldots, A_n^k$. Conversely, the firing strength of the other rules in the rule base will be low since they represent other regions of the input space. In this case, the output of the model will strongly be dominated by $g^k(\cdot)$ the consequent function which models the relation between $\underline{X}$ and Y in region k defined by the premise of R^k. The ability to partition the input space allows this type of model to represent even highly non-linear relations between $\underline{X}$ and Y. Furthermore, since the partition of the input space is fuzzy, the model can represent the relation for inputs located near the borders of the regions. This is performed by computing the weighted average of the consequent responses $g^i(\cdot)$ of rules which model "contiguous" regions of the input space.

We have defined the general structure of the fuzzy model used in the F-NN forecasting model. However, the type of membership functions used to describe the fuzzy sets in the premises of the rules, as well the type of function in the consequent part of the rules need to be defined. As noted in [46], several types of membership functions can be used: triangular, trapezoidal, Normal, etc. In the present case, the membership functions used are Normal functions, hence:

$$\psi_{A_j^i}(x_j) = \exp\left(-\left(\frac{x_j - a_j^i}{b_j^i}\right)^2\right), \tag{4.17}$$

where a_j^i and b_j^i respectively represent the two parameters of the Normal function. With respect to the consequent function, several possibilities are also available, two are described in [46]: a constant and a linear function of the input variables. Here, we will only consider the constant function:

$$g^i(\underline{x}) = p^i. \tag{4.18}$$

From this, the analytical expression of the fuzzy model can be derived by introducing Equation 4.17 and Equation 4.18 into Equation 4.16. The model can therefore be expressed as:

$$f(\underline{x}) = \frac{\sum_{i=1}^{m} p^i \prod_{j=1}^{n} \exp\left(-\left(\frac{x_j - a_j^i}{b_j^i}\right)^2\right)}{\sum_{i=1}^{m} \prod_{j=1}^{n} \exp\left(-\left(\frac{x_j - a_j^i}{b_j^i}\right)^2\right)} \tag{4.19}$$

We have defined the analytical expression of the fuzzy-neural network model. However, to use this model to compute forecasts two aspects must be addressed: model parameter identification and structure or architecture identification. Parameter identification refers to the estimation of the values of the a_j^i, b_j^i, and p^i parameters in Equation 4.19. Architecture identification refers to defining the number of fuzzy rules that should be used to model a process. This step is equivalent to the search for the best topology of a neural-network, where the number of hidden neuron layers and the number of neurons per layer must be defined. The way these aspects are carried out in Armines Wind Power Prediction System (AWPPS) is described in the following sub-sections.

Parameter Identification

The aim of parameter identification, or model training, is to determine the values of the model parameters. The model parameters are adjusted using a stochastic gradient decent method. In this method the model parameters are adjusted by minimizing a cost function E_N computed over a training data set. Here, E_N is defined as:

$$E_N = \frac{1}{N} \sum_{t=1}^{N} E(t) \tag{4.20}$$

where N is the number of observations in the training set and $E(t)$ a cost function defined for time t. The cost function $E(t)$ is defined as:

$$E(t) = \frac{1}{2} e(t)^2 \tag{4.21}$$

where $e(t)$ is the model error at time t. The model error is $e(t) = y(t) - \hat{y}(t)$, where $y(t)$ is the observation at time t and $\hat{y}(t)$ is the model output associated to the model input $x(t)$ for time t.

The stochastic gradient descent method is employed to derive the learning rules necessary to estimate the model parameters. This method allows the minimisation of the total learning error E_N through the minimization of the instantaneous error $E(t)$. To determine the model parameters, the cost function $E(t)$ is differentiated with respect to each parameter. Let $z = [z_1, z_2, \ldots, z_q]^T$ be the vector of adjustable model parameters (in our case these are: a_j^i, b_j^i, and p^i), the derivative of $E(t)$ with respect to z_i is called the gradient of the error surface and noted $\nabla_{z_i} E(t)$ with $i = 1, \ldots, q$:

$$\nabla_{z_i} E(t) = \frac{\partial E(t)}{\partial z_i} \tag{4.22}$$

The parameter values are found according to the steepest decent method where the parameters are adjusted along the error surface in order to converge towards the optimal solution. The parameters are adjusted in the direction of the steepest error descent which is the opposite of the gradient $\nabla_{z_i} E(t)$. The parameter variation is then:

$$\Delta z_i = -\eta \nabla_{z_i} E(t) \Rightarrow z_i(t+1) = z_i(t) - \eta \frac{\partial E(t)}{\partial z_i} \tag{4.23}$$

where η denotes a learning rate parameter. This parameter is variable through time, that is through the model learning epochs. Its evolution is governed by a learning rate schedule described in [46]. Note that in the present model three types of parameters are used: a, b and p which stand for the a_j^i, b_j^i, and p^i parameters. For each type of parameter a different learning rate schedule, and hence a different learning rate η_z is used.

After some algebra the learning rules are obtained. The learning rules can be written as:

$$\begin{aligned}
a_j^i(t+1) &= a_j^i(t) + \eta_a \cdot \frac{2\left(x_j - a_j^i(t)\right)}{b_j^i(t)^2} \cdot \tilde{w}^i \cdot \left(p^i - \hat{y}\right) \cdot e(t) \\
b_j^i(t+1) &= b_j^i(t) + \eta_b \cdot \frac{2\left(x_j - a_j^i(t)\right)^2}{b_j^i(t)^3} \cdot \tilde{w}^i \cdot \left(p^i - \hat{y}\right) \cdot e(t) \\
p^i(t+1) &= p^i(t) + \eta_p \cdot \tilde{w}^i \cdot e(t)
\end{aligned} \tag{4.24}$$

where p^i is the constant output of rule i as defined in Equation 4.18 and $\tilde{w}^i$ is the normalized weight of each rule:

$$\tilde{w}^i = \frac{w^i}{\sum_{i=1}^{m} w^i} \tag{4.25}$$

Architecture Identification

The second aspect that must be addressed when building the fuzzy-neural network model is its architecture. This translates to defining the number of fuzzy rules necessary to appropriately model the process under consideration. In the present case, the architecture of the model is defined by the number of fuzzy sets that are associated to the universe of discourse of each input variable. Once the number of fuzzy sets for each variable is defined the rule base can be generated. To illustrate the way in which the rule base is generated let us consider a case with two input variables X_1 and X_2. Let us also assume that two fuzzy sets are associated to each variable: sets $A_{1,1}$ and $A_{1,2}$ are associated to variable X_1 and sets $A_{2,1}$ and $A_{2,2}$ are associated to variable X_2. The rule base is derived as follows:

$$\begin{aligned}
R^1 &: \text{IF } x_1 \text{ is } A_{1,1} \text{ and } x_2 \text{ is } A_{2,1} \text{ THEN } p^1 \rightarrow R^1 : \text{IF } x_1 \text{ is } A_1^1 \text{ and } x_2 \text{ is } A_2^1 \text{ THEN } p^1 \\
R^2 &: \text{IF } x_1 \text{ is } A_{1,1} \text{ and } x_2 \text{ is } A_{2,2} \text{ THEN } p^2 \rightarrow R^2 : \text{IF } x_1 \text{ is } A_1^2 \text{ and } x_2 \text{ is } A_2^2 \text{ THEN } p^2 \\
R^3 &: \text{IF } x_1 \text{ is } A_{1,2} \text{ and } x_2 \text{ is } A_{2,1} \text{ THEN } p^3 \rightarrow R^3 : \text{IF } x_1 \text{ is } A_1^3 \text{ and } x_2 \text{ is } A_2^3 \text{ THEN } p^3 \\
R^4 &: \text{IF } x_1 \text{ is } A_{1,2} \text{ and } x_2 \text{ is } A_{2,2} \text{ THEN } p^4 \rightarrow R^4 : \text{IF } x_1 \text{ is } A_1^4 \text{ and } x_2 \text{ is } A_2^4 \text{ THEN } p^4
\end{aligned}$$

The rules are derived by defining all the two-fuzzy-set combinations of the fuzzy sets initially defined for X_1 and X_2. Then, the number of fuzzy sets is doubled in order to obtain

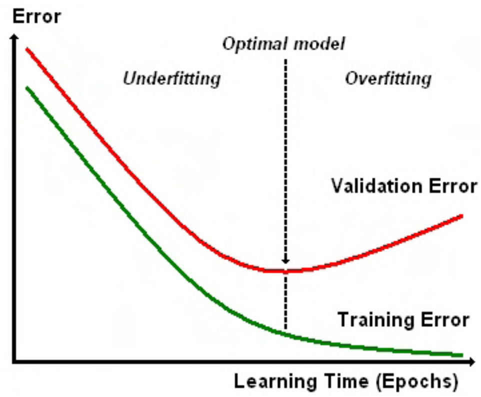


FIGURE 4.2: *Typical evolution of model error during learning on the training and validation data sets*

fuzzy sets that are independent between rules. The number of rules is thus the product of the number of fuzzy sets associated to each input variable.

In this way the architecture of the model can be defined as a vector where each element represents the number of initial fuzzy sets associated with each input variable. In the present case, additional elements are added to the architecture vector. These elements represent initialisation values for the learning rate schedule algorithms governing the evolution of the learning rates for each type of model parameter a, b and p.

The aim of architecture definition is to ensure that the resulting model is capable of capturing the statistically significant properties of the process being modelled, while at the same time avoiding overfitting. In the present case, a model with a given architecture is trained using cross-validation to avoid overfitting. In the cross-validation scheme the generalisation capability of a model can be evaluated during the learning phase. To do this, the learning data set is divided into a training set and a validation set. The model tunes its parameters by examining the training set several times or epochs. After each training epoch, performed on the training set, the model's parameters are fixed and its performance is tested on the previously "unseen" validation set. In this way the performance of the model is computed for each epoch on both data sets. At the end of the learning phase the modeller can estimate the generalisation capability of the model by comparing the evolution of the error on the training set and the validation set. As can be seen in Figure 4.2, at the beginning of the learning time the error on the validation set is quite high. As learning progresses, the error on both data sets decreases, and the error on the training set decreases to very low values. However, after a number of learning epochs, the error on the validation set may increase. This denotes overfitting of the model. Intuitively, the model has started learning the noise in the training dataset, which leads to higher error on the validation set. To obtain the best model parameter estimation, early-stopping is applied. The model training is stopped when the error reaches the minimum value on the validation set.

In order to compare the performance of different model architectures, the best performance obtained on the validation set for each architecture are compared. The architecture that achieves the lowest error is considered to be the one that leads to the best generalisation. Defining a good architecture is often done by a trial-and-error that requires problem specific expertise as well as considerable engineering time. In the present work, an automated approach based on an optimisation algorithm is used. Here, the vector describing a model's architecture can be considered as belonging to a hyperspace of all possible model architectures. In the frame of optimisation, the model error associated to the architecture vector becomes a solution space that the optimization algorithm must explore in order to find a good solution. Given the non-linear nature of the optimization problem [46], a non-linear Simplex method named the "Complex" method proposed by Box is used. This method is a sequential search technique that explores the solution space by evaluating the curvature of the objective function through the evaluation of a set of solutions, which form a geometric figure called a "simplex". In [46] a variant is proposed, capable of treating mixed integer and continuous optimization problems. The method is described in detail in [46], here only a general description of the method is presented.

The idea behind the complex algorithm is to evaluate the objective function for a small number of possible initial solutions (in our case 24 points). Once the value for each point of the complex is known, the point, or architecture, for which the model error is highest, is "reflected" toward the centroid of the complex. This means that the coordinates of the point are modified so that a new point located closer to the centroid is defined. The value of the objective function is evaluated for this new point. If the error value is lesser than that found for the initial point, this point replaces the previous one in the complex and the procedure is repeated for the point in the complex which has the highest error. If the error at the reflected point is not inferior to that of the initial point, another reflection is carried out in order to evaluate a new point closer to the centroid. In this way the reflected point is progressively moved toward the centroid. After a given number of "failed" reflexions the point is replaced in the complex by the centroid and the procedure is started over. Every time a valid reflection is obtained the convergence of the algorithm is checked. The algorithm is considered as having converged if for a certain number of iterations the difference between the minimum and maximum values for the simplex points are below a certain, problem dependent, threshold.

Wind power forecasting with the fuzzy neural network

The fuzzy model we have described is a generic prediction model. It can be used to forecast time series in a classical auto-regressive manner, but can also forecast time series using exogenous explanatory variables. In the present context, the aim is to use both on-line power measurements and NWP forecasts in order to provide wind power forecasts for the next 48 to 72 hours, with hourly time steps. In order to compute multi-step ahead predictions the model is used in a recursive way. That is, for the first forecast step the model can for example be written as:

$$\hat{P}(t+1\,|\,t) = f(P(t), \hat{W}s(t+1\,|\,t), \hat{W}d(t+1\,|\,t)) \tag{4.26}$$

where $\hat{P}(t+1\,|\,t)$ is the forecast for time $t+1$, $P(t)$ is the power measurement for time t, and $\hat{W}s(t+1\,|\,t)$, $\hat{W}d(t+1\,|\,t)$ are respectively the NWP wind speed and wind direction forecasts for time $t+1$, and f is the fuzzy-neural network. For horizon $t+1$, the power measurement at time t is used as input. For the following forecast steps no measurements are available, so the power forecast computed for the previous hour is used instead. The model for the following steps can be written:

$$\hat{P}(t+k\,|\,t) = f(\hat{P}(t+k-1\,|\,t), \hat{W}s(t+k\,|\,t), \hat{W}d(t+k\,|\,t)) \tag{4.27}$$

where $\hat{P}(t+k\,|\,t)$ is the power forecast computed for horizon $t+k$ and $\hat{P}(t+k-1\,|\,t)$ is the power forecast computed for the previous horizon $t+k-1$.

During the cross-validation process, the model error over the validation set is computed over all forecast horizons, in this way the error over all horizons is minimized. This is valid for the model architecture optimization phase as well as for the final training of the model.

Below we discuss this model in the wind power forecasting context. We can start by mentioning the models advantages. The model is very generic, it can be used for many forecasting purposes, its use is not restricted to wind power forecasting. Because it is generic, the model can be run with any relevant input data. In the frame of wind power forecasting this means that a wealth of explanatory variables such as wind direction, air temperature, atmospheric pressure, and others, can be used to improve forecast accuracy. With this in mind, the architecture optimization can, to some extent, select input data. Indeed if the "optimal" architecture has zero fuzzy set associated to an input variable then this variable is not considered by the model. The automatic architecture optimization can also free engineer time in an operational setting where many models must be set-up for an end-user. The optimization usually leads to models having good forecasting performance on previously unseen data without any modeller intervention. Nonetheless, the model does have some disadvantages. These disadvantages stem in part from its advantages. The generic nature of the model requires that many parameters be defined: the number of fuzzy sets for each variable, the learning rate schedule initialization values. Although these parameters are "optimized" by the Complex algorithm, the algorithm also requires some parameter initialization, which is left to the experience of the modeller. For example, the modeller must define the parameters of the complex reflection procedure. Furthermore, although of an automatic nature, the architecture optimization phase is rather time consuming. Typical optimization times are in the order one week to ten days. Although such computation times are possibly acceptable in an operational setting, where the best possible performance is expected, this can be hampering in a research perspective where, often, many case studies must be run in order to study some aspect of the forecasting problem.

4.4.2 Regressive Power Curve Model

In order to compare the performance of the F-NN model on the regional forecasting problem to that of another model, we have chosen to develop another statistical model based on a very different modelling approach. As mentioned above, the fuzzy model is very generic it can be used to forecast any type of time series. This is clearly shown by its performance on some classic benchmark cases such as the sunspot series and the Makey-Glass series [46] as well as in short-term wind power forecasting in many different terrain and meteorological conditions [32, 64]. This multi-usefulness comes at the price of a rather lengthy optimization and tuning process.

The primary property of the model proposed here should be fast model optimization and tuning. The objective is to define a model for which off-line performance values can rapidly be obtained. Of course, the model should have acceptable performance. By "acceptable" we mean performance values, in terms of the criteria defined in [34], that are within the range of performance of other models found in the literature. Because of these requirements, the model need not be totally generic.

To answer these requirements, we propose the model presented in the next paragraphs that we call Regressive Power Curve (RPC) model. This model is based on a simple statistical power curve modelling-scheme and its optimization and tuning times are much shorter than those of the F-NN model.

Defining the Regressive Power Curve Model

As suggested by the data characterization in subsection 3.6.4, the variable provided by NWP forecasts that seems to be the most useful in wind power forecasting is wind speed. Indeed, when compared to wind direction, this variable clearly contains the most information on the future regional power generation. This is true for the cases presented in section 3.5, and is also valid in most single wind farm cases. Because of this, the RPC model is designed to specifically model the relation between predicted wind speed and future power generation. The model is not designed to take into account any other NWP variable. However, as we will see in section subsection 4.6.1, neglecting wind direction, and other meteorological variables, can sometimes lead to decreased forecast accuracy when compared to models such as the F-NN model.

The RPC model is based on a transformation of the NWP wind speed forecast values into power using an approximation of the characteristic curve of the wind farm. A simplest way of approximating the power curve would be to use the wind turbines manufacturer's power curve and the wind speed forecast extrapolated to the hub height using the logarithmic profile of the wind. Although simple, this approach may provide reasonably good forecasts. However, the sum of the wind turbine power curves is not necessarily a good model for the wind farm. Furthermore, when NWP forecasts are for a grid of points around the wind farm, a downscaling procedure would have to be applied to have reliable wind speed forecasts at the level of the wind farm. This would require a significant amount of static information on the wind farm (i.e. terrain roughness, orography, etc), which might no be readily available.

In order to obtain a model that can reduce the above-mentioned approximations and whose only information requirements are the time series under consideration, a statistical approach was preferred. The aim here is to model the relationship between the wind speed forecasts and the power output of the wind farm without any other considerations.

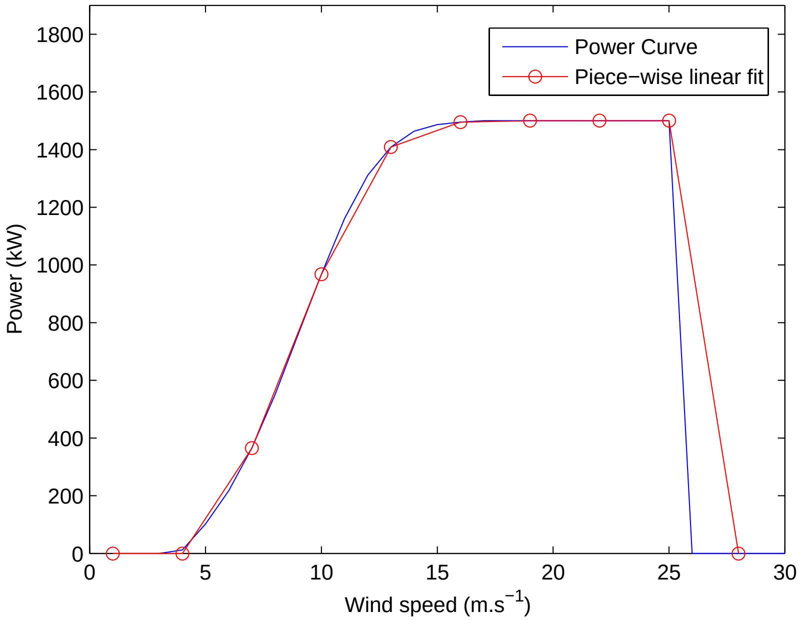


FIGURE 4.3: *Piecewise linear fitting of a power curve*

To find the relation between one variable and another, many different types of fitting can be used. Artificial intelligence models are one possibility, however this would lead to a model having computational requirements similar to the F-NN model. Other types of fitting pose similar problems such as polynomial fitting, where the degree of the polynomial must be optimised. In order to keep computation time short, while at the same time providing sufficient modelling power, linear fitting is preferred. To model the power curve, a piecewise least squares linear fitting of the wind-speed to power relation is proposed. As can be seen in Figure 4.3 the principle of piecewise linear fitting is to divide the domain of the explanatory variable, in our case wind speed, into several bins. For each bin a linear function is fitted to the modelled function. Here, the model $f(Ws)$ of the power curve can be written as:

$$f(ws) = \begin{cases} \alpha_1 Ws + \beta_1 & \text{if } \gamma_0 \leq Ws < \gamma_1 \\ & \vdots \\ \alpha_i Ws + \beta_i & \text{if } \gamma_{i-1} \leq Ws < \gamma_i \\ & \vdots \\ \alpha_m Ws + \beta_m & \text{if } \gamma_{m-1} \leq Ws < \gamma_m \end{cases} \tag{4.28}$$

where Ws is the wind speed, α_i and β_i are the parameters of the local linear approximation of the power curve corresponding to bin i and $[\gamma_{i-1}, \gamma_i[$ are the m bins over which the least squares fitting is performed.

The above model is the basis for the proposed RPC model. From Equation 4.28 it is clear that the size of the bins used to compute the local approximations will have an impact on the performance of the model. Very small bins will capture the noise in the data and very large bins will lead to an over-smooth approximation of the power curve.

Also, as shown in the data characterization presented in subsection 3.6.4, the amount of information that wind speed forecasts provide for different horizons varies. Hence, the relation between forecast wind speed and observed power production is horizon-dependent. To account for the variation of the relation with respect to horizon, a separate power curve is defined for each forecast horizon. In this way, variations in the NWP performance for different horizons can be captured and corrected to some extent. The obtained predictive power curve can thus be written as a function of the predicted wind speed $\hat{W}s(t+k\,|\,t)$ and horizon k as:

$$f_k(\hat{W}s(t+k\,|\,t)) = \begin{cases} \alpha_{1,k}\hat{W}s(t+k\,|\,t) + \beta_{1,k} & \text{if } \gamma_{0,k} \leq \hat{W}s(t+k\,|\,t) < \gamma_{1,k} \\ & \vdots \\ \alpha_{i,k}\hat{W}s(t+k\,|\,t) + \beta_{i,k} & \text{if } \gamma_{i-1,k} \leq \hat{W}s(t+k\,|\,t) < \gamma_{i,k} \\ & \vdots \\ \alpha_{m,k}\hat{W}s(t+k\,|\,t) + \beta_{m,k} & \text{if } \gamma_{m-1,k} \leq \hat{W}s(t+k\,|\,t) < \gamma_{m,k} \end{cases} \tag{4.29}$$

By using the model described above, the RPC model can convert the forecast wind speed values into future wind power production values. As such, the model can be expected to have good performance as long as the meteorological forecasts are sufficiently accurate. It is however remarked that the performance of NWP forecasts is often best for horizons over six hours ahead. Furthermore, it is widely known that if on-line power measures are available through a SCADA system, they can contribute to improve the performance of a prediction model for short-term horizons (i.e. 0-6 hours). Hence, to potentially improve the short-term performance of the piecewise model, including on-line power measurements in the final prediction is necessary. To take into account power measurements, a simple linear combination of the power curve model output and the last available power measure is proposed. Because the power curve modelling is horizon-dependent, and the "usefulness" of the power measurement decreases with the horizon, a separate linear relation is defined for each horizon. In this way, the output for each forecast horizon k can be written as:

$$F_k(\hat{W}s(t+k\,|\,t), P(t)) = a_k f_k(\hat{W}s(t+k\,|\,t)) + b_k P(t) + c_k \tag{4.30}$$

where $F_k(\cdot)$ is the RPC model overall function, $P(t)$ is the power measured at time t, f_k is the power curve computed for horizon k, and a_k, b_k, and c_k are the linear combination coefficients.

Parameter Identification

As with the fuzzy-neural network, the parameters of the RPC model must be identified. The identification is performed by a "learning" process based on a learning data set. As can be seen from Equation 4.29 and Equation 4.30, the RPC model combines two linear models in two successive steps. Given that both model steps are linear and that one of the objectives of the model is to possess fast learning capabilities, least-squares linear regression was retained as the parameter estimation method.

Here, a rapid description of least-squares linear regression is given. The aim of the method is to minimize the sum of square errors. Let the general regression equation be:

$$y_i = a_1 + a_1 x_{i,2} + a_2 x_{i,3} + \ldots + a_k x_{i,k} + e_i \tag{4.31}$$

where:

- y_i is the ith observation of the modelled process in the learning set,
- $x_{i,2}, \ldots, x_{i,k}$ are the explanatory variable values associated to observation i,
- e_i is the model error for the ith observation and,
- $a_1, \ldots, a_k$ are the model parameters to be estimated.

For all observations in the learning set a matrix expression can be used:

$$\mathbf{Y} = \mathbf{X}\,\mathbf{a} + \mathbf{e} \tag{4.32}$$

where:

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{1,2} & x_{1,3} & \ldots & x_{1,q} \\ 1 & x_{2,2} & x_{2,3} & \ldots & x_{2,q} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{N,2} & x_{N,3} & \ldots & x_{N,q} \end{bmatrix}, \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_q \end{bmatrix}, \text{and } \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}.$$

To obtain the value of $\mathbf{a}$, the sum of square errors must be minimized:

$$SSE = \sum_{i=1}^{N} = e_i^2 = \mathbf{e}'\,\mathbf{e} \tag{4.33}$$

where $\mathbf{e}' = (\mathbf{Y} - \mathbf{X}\,\mathbf{a})'$, is the transpose of $\mathbf{e}'$. The problem of minimizing the sum of square error is solved by taking partial derivatives of SSE with respect to each coefficient a_i, setting these derivatives equal to zero, and solving a set of k equations in k unknowns to obtain estimates of the a_i parameters.

In matrix notation, the partial derivative of SSE set equal to zero is [40]:

$$\frac{\partial \mathbf{e}'\mathbf{e}}{\partial \mathbf{a}} = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\mathbf{a} = 0, \tag{4.34}$$

from this

$$\mathbf{X}'\mathbf{Y} = \mathbf{X}'\mathbf{X}\mathbf{a} \tag{4.35}$$

and

$$\mathbf{a} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{4.36}$$

where $(\mathbf{X}'\mathbf{X})^{-1}$ denotes the inverse of matrix $(\mathbf{X}'\mathbf{X})$.

The linear regression parameter identification is achieved by applying Equation 4.36 to the learning data set for the two sub-models that constitute the RPC model. To apply this to the power curve model f_k, the learning data points are sorted into the bins defined by $[\gamma_{i-1,k}, \gamma_{i,k}[$, then for each bin, the $\alpha_{i,k}$ and $\beta_{i,k}$ parameters are computed by solving

$$\mathbf{a}_{i,k} = (\mathbf{X}'_{i,k}\mathbf{X}_{i,k})^{-1}\mathbf{X}'_{i,k}\mathbf{Y}_{i,k} \tag{4.37}$$

where $\mathbf{a}_{i,k} = (\alpha_{i,k},\, \beta_{i,k})'$, $\mathbf{X}_{i,k}$ are the predicted wind speeds for horizon k and $\mathbf{Y}_{i,k}$ are the associated production values.

Once the piecewise function parameters have been computed for all horizons, the power curve forecasts $f_k(\hat{ws}(t+k\,|\,t))$ are computed for all observations in the learning set. The parameters of the combination model used to merge the power curve outputs and the power measurements can be computed using Equation 4.36. In this case:

- $\mathbf{a}_{i,k} = (a_k,\, b_k,\, c_k)'$ and,
- $\mathbf{X}_k$ is the explanatory variable matrix containing the power forecasts derived from the power curve models as well as the past production variables, and $\mathbf{Y}_k$ is the associated observation matrix.

Architecture Identification

Up to now the number and size of bins used in the modelling of the power curve for a given horizon has been considered as predefined. As mentioned previously, the number and size of the bins will directly influence the performance of the model. Adequately defining the bins used in the RPC model can be compared to defining the F-NN model architecture. In this section we will discuss the choice of the number and the size of these bins.

The first aspect that must be considered is the number of bins. Indeed, this will directly influence the number of parameters in the model and, by extension, the modelling capacity of the model. Intuitively, this can be understood by considering two extreme cases. In the first case let us consider that only one bin is defined. Hence the power curve, which has a characteristic sigmoid shape followed by null values after the cut-off threshold, is modelled by a straight line. This will necessarily lead to high forecast errors. The second case can be one where the number of bins is equal to the number of possible wind speed values provided by the NWP model; typically NWP models provide wind speed values with 0.1 m.s^{1} precision. In this case one specific value would be modelled for each possible input value. Hence, considering that observations are available for each possible input variable, the model will capture much of the noise in the learning data, which will lead to bad generalization due to overfitting. Hence, the number of bins must be "optimized" in order to allow sufficiently fine modelling of the process while at the same time minimizing the risk of overfitting.

The second problem that must be addressed when determining the bins is their size or the location of the bin bounds. The problem can be clearly illustrated by examining the theoretical fitting presented in Figure 4.3. In this figure, 10 equal size bins are used to model the power curve. The power curve plateau located between 16 m.s^{-1} and 25 m.s^{-1}, is modelled by three linear functions. However, since the power curve is constant between 16 m.s^{-1} and 25 m.s^{-1}, it can be modelled just as precisely by a single linear function defined on the $[16\,\mathrm{m.s}^{-1},\,25\,\mathrm{m.s}^{-1}[$ bin. In this case, the two functions that are no longer necessary to model the plateau can be used to better model parts of the power curve presenting high second derivatives. In this way the model error can be reduced without increasing the number of functions used to model the power curve.

The third problem that must be addressed is the number of observations that fall into each bin. Depending on the bin definitions, some bins might contain too few observations to compute model parameters that adequately reflect the relation between the explanatory and the dependent variable. Therefore the distribution of the explanatory variable must be taken into account when defining the bins. In the frame of wind power forecasting, the wind speed forecast values follow a distribution, which follows, more or less closely, the true wind speed distribution. Because of this, high forecast wind speed values are rather rare. Hence, depending on the length of the learning set, defining small bins for wind speed values above say 20 to 22 m.s^{-1} can lead to bins for which few or no observations are available. In this case, the model parameters can be highly biased or it can be impossible to compute them. This, in turn, can lead to highly inaccurate predictions when the model is tested on previously unseen data.

Several possibilities exist to define the bins. One possibility is to rely on the experience of the modeller. This would be similar to membership function definition in a fuzzy expert system. This however would entail a manual trial and error procedure where the modeller progressively refines the bin definition. Given that RPC model associates one power curve to each forecast horizon, this irksome procedure would have to be repeated for each horizon, typically 48 to 72 times. Clearly this is not the most convenient way to "optimize" the bins. Another possibility is to use an optimization algorithm that tests the performance of different bin definitions following a cross-validation procedure similar to the one used for the F-NN model. Indeed, the RPC model can be trained on a training set with different bin definitions. The performance on a validation set can be used to determine the bin definition most likely to lead to the best generalisation. Such an automatic procedure is clearly more convenient and likely to lead to better results than a trial and error method. Therefore, this is the method retained to define the bins in the present work.

The decision variables of the optimization problem at hand must be defined. In the present case, we opt for the number of observations that fall into each bin. The objective is to define the bins so that each bin contains an equal number of observations. This type of binning, known as equal mass binning, ensures that each bin contains enough observations to obtain a reasonably accurate estimation of the piecewise function parameters. This choice effectively deals with the problem of defining the number of observations in each bin. Furthermore, this automatically sets the number of bins that will be used depending on the length of the learning set. Finally, the definition of the bin bounds is straightforward since the problem is reduced to sorting the explanatory variable observation and defining the bounds so that each bin contains the same number of observations. We must note here that given the precision of the values provided by the NWP models, the bins will not contain exactly the same number of observations. Therefore the aim is for each bin to contain a minimum, or target, number of observations. In the present case the bin bounds are defined as follows: the input variable is sorted in ascending order. The lower bound of the first bin is set equal to the lowest possible value. The upper bound is the value for which the bin contains at least the target number of observations. This bound constitutes the lower bound of the following bin. The upper bound of the following bin is again set as the value for which the next bin contains at least the target number of observations.

To solve this optimization problem, the Complex algorithm used for the fuzzy-neural network could be used. However, here we only have one decision variable and the only constraint is that the target number of observations be inferior to the number of observations in the training set. Also, given that one “architecture” can be evaluated very rapidly, a simpler algorithm (Algorithm 4.1) that conducts a more complete search of the solution space was preferred.

Algorithm 4.1 RPC bin definition algorithm

for Horizon $k = 1$ to $maxHorizon$ **do**
  for Target number of observations $nbObs = minNbObs$ to $maxNbObs$ **do**
    Compute the $\alpha_{i,k}$, $\beta_{i,k}$, a_k, b_k, and c_k parameters on the learning set according to the bin definition derived from $nbObs$
    Using these parameters, compute the sum of square errors on the validation set.
    Store the validation sum of square errors.
  end for
  Find $nbObs_k^*$ the “optimal” target number of observations for horizon k
end for

This algorithm searches all possible architectures and selects the one that leads to the best performance on the validation set. The only initialisation parameter is the minimum target number of observation. This should be set so that the estimation of the parameters leads to statistically significant values.

The algorithm can be refined depending on the sensitivity of the sum of square errors to the target number of observations. In the case of wind power forecasts it was found that the performance on the validation set was relatively insensitive to small variations in the target number of observations. Hence, examining all possible values is not always necessary. The target number can be increased from one iteration to the next by more than one. In this way, fewer architectures are tested, which speeds up the optimisation process. In the results presented later this property was exploited, the target number increase between two architectures ranged from 5 to 10 observations.

4.5 Scope of the study

The aim of the study conducted here is to compare the performance of different models used in different regional forecasting configurations. Here, the performance of the F-NN and RPC models is compared for the three basic configurations described in section 4.3, namely the direct forecasting scheme, the cascaded scheme and the cluster scheme. The paradigm of the Anemos competition [32] is followed in this investigation. The aim is to detect difference in performances that can be linked either to the models themselves, to the modelling approach, or to the complexity and the inherent predictability of the different cases studies.

In order to perform the comparison between the two models the first step is to validate the performance of the RPC model. Indeed, although other state-of-the-art models inspire this model, it is designed on much simpler principles, hence its performance must be tested. Our aim being to examine the behaviour of advanced models in a regional forecasting setting, we must first ensure that the RPC performs well when compared to other state-of-the-art models. To perform this validation, the ideal case would have been to possess a data set on which other regional forecasting models have previously been evaluated. Such a data set was unfortunately unavailable. However, the datasets used for the Anemos single wind farm forecasting competition were available. This permits to compare the performance of the RPC model to that reported for other state-of-the-art models benchmarked in this competition. In this way, the performance of the RPC model for single wind farm forecasting can be evaluated, and its use as a reasonable base-model in the regional forecasting setting can be justified.

In a next stage, for the regional forecasting comparison, the F-NN and RPC models are evaluated on the Irish and Danish case studies described in section 3.5. In order to compare the results, the evaluation methodology proposed in [32] was followed here. For each test case the same NWP data was available to both models. For all cases a common learning, validation and testing set was defined. The final comparison is conducted on the results of the testing set. The performance of the models is measured using the criteria defined in section 2.3.

4.5.1 Validation of the RPC model for single wind farm forecasting

To validate this model as a good base-model for regional forecasting evaluation purposes, the model was evaluated on the single wind farm forecasting problem. The evaluation was done on the datasets used in [32]. More specifically, data were used from the Tunø Knob, Golagh, and Alaiz wind farms in Denmark, Ireland and Spain respectively.

The Tunø Knob wind farm is an offshore farm located 6 km off the east coast of the Jutland peninsula and 10 km west of Samsø island. The wind farm has a rated capacity of 5 MW. The available data cover a period from March 18, 2002 to April 30, 2003. The available NWP data is provided by the Hirlam 0.15° model operated by the Danish meteorological service. The forecasts are updated four times a day with a forecast horizon of 48 hours with hourly time steps. The NWP data and production data are split into a learning and a validation set. The learning set spans the period from March 18, 2002 to December 15, 2002. The testing set runs from December 16, 2002 to April 30, 2003.

The Golagh wind farm is an on-shore farm located in county Donegal in northwestern Ireland. The wind farm has a rated capacity of 15 MW. The available data cover the period from August 1, 2002 to March 31, 2003. The available NWP data are provided by the Hirlam model operated by the Irish meteorological service. The NWP data are updated four times a day with a forecast horizon of 48 hours with hourly time steps. The learning set spans the period from August 1, 2002 to January 31, 2003. The testing set runs from February 1, 2003 to March 31, 2003.

The Alaiz wind farm is located 15 km south of Pamplona in the Navarre community, in the northwestern part of Spain. The farm is located in very complex mountainous terrain. The farm is located between 910 and 1120 meters above sea level, and has a rated capacity of 33 MW. The available data cover all of 2001. The NWP data are provided by the Hirlam 0.2° model operated by the Spanish meteorological service. The NWP forecasts are updated four times a day with a forecast horizon of 24 hours with hourly time steps. The learning set spans the period from January 1, 2001 to August 31, 2001. The testing set runs from September 1, 2001 to December 31, 2001.

4.5.2 Regional Forecasting Approaches

To examine the performance of different modelling approaches, the Danish and Irish data sets were used. Because the Danish data were provided at a later stage of this work, some analysis only the Irish case was evaluated.

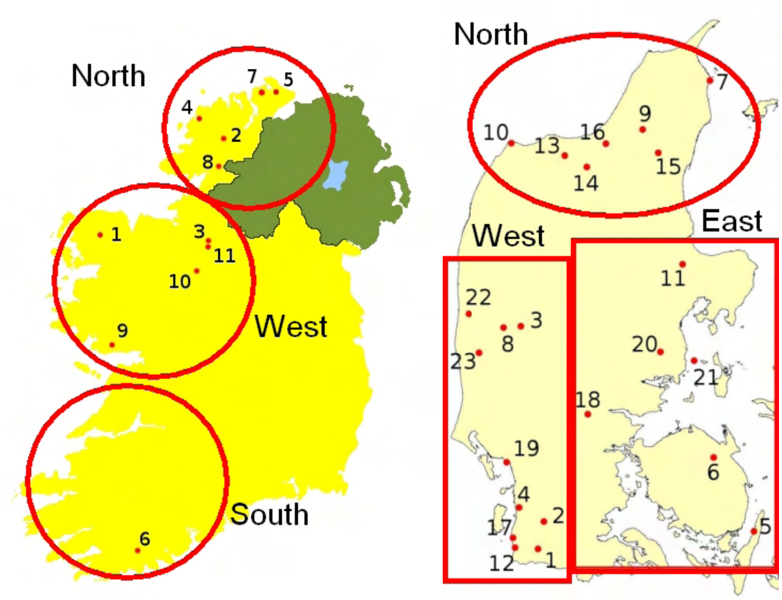
When using the Danish dataset, the available data were split into a learning and testing sets as follows: the first 9 months for learning (December 31, 2002.12 to September 30, 2003) and the remaining period (October 1, 2003 to July 22, 2004) for testing and model performance evaluation. This data partition was used for all model evaluations.

When using the Irish dataset, the available data were split into a learning and testing set as follows: the first 10 months for learning (February 5, 2001 to December 18, 2001) and the remaining period (December 19, 2001 to July 15, 2002) for testing. This data partition was used for all model evaluations.

Direct Approach

We remind the reader that the direct approach uses the statistical model in a single step to compute the regional forecast. No intermediate forecasts are computed. To examine this approach, two studies were conducted. In the first one the data from a single reference wind farm was used as input to the forecasting models. This evaluation was carried out on the Irish data set. Wind farm 8 was used as the reference farm.

In the second study, averaged NWP data and the sum of available single wind farm productions was used as input. This study was carried out on the Danish case study, the averaged NWP data was computed using the data from the 23 reference wind farms. The on-line



(a) Ireland (b) Denmark

FIGURE 4.4: *Wind farm clusters defined for the Irish and Danish case studies.*

production data is the sum of the 23 reference farm productions and the forecast production is that of the total Jutland-Funen region.

Cascaded Approach

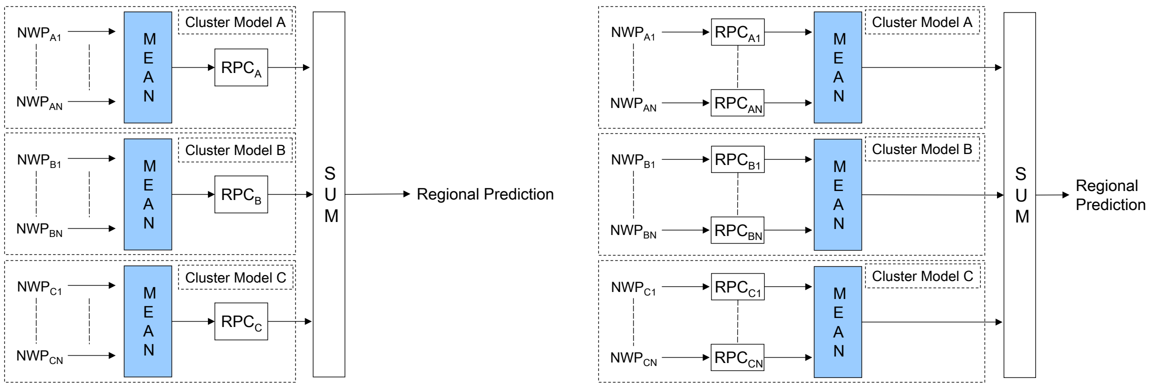
In the cascaded approach individual wind farm forecasts are used as input to a regional forecasting model. Here, the Irish dataset was used to compare the F-NN and RPC model performance. For each of the eleven reference wind farms, an F-NN and RPC model was tuned to provide forecasts of the farm's future production. The individual forecasts were then summed in order to obtain a forecast of the total regional production.

Cluster Approach

In the cluster approach the wind farms in a region are divided into several clusters. The aim is to compute production forecasts for each cluster and then sum the cluster forecasts to obtain the regional production forecasts. To evaluate the F-NN and RPC models in this approach, the Irish case was used. The wind farms were grouped into three clusters as shown in Figure 4.4(a). For each cluster, the NWP forecasts provided for each wind farm in the cluster were used as input to the forecasting models. In reality only wind farm 8 possesses a reliable SCADA system. Because of this, only the northern cluster model was supplied with on-line information from wind farm 8. Because wind farm 6 is the only wind farm in the southern cluster, a single wind farm forecasting approach was followed to compute the forecasts for this cluster.

4.5.3 Comparison of modelling approaches

The modelling approaches described above are expected to lead to varying results depending on the input data used by the forecasting models. Comparing different modelling approaches with the same input data is necessary in order to establish the impact each approach may have on the forecast performance. To compare two modelling approaches with the same input, two cluster approaches will be compared on the Irish and Danish case studies. The cluster definition for the Danish case can be seen in Figure 4.4(b).



(a) Mean NWP (b) Mean Cluster

FIGURE 4.5: *Two cluster forecasting approaches exploiting the same on-line input data.*

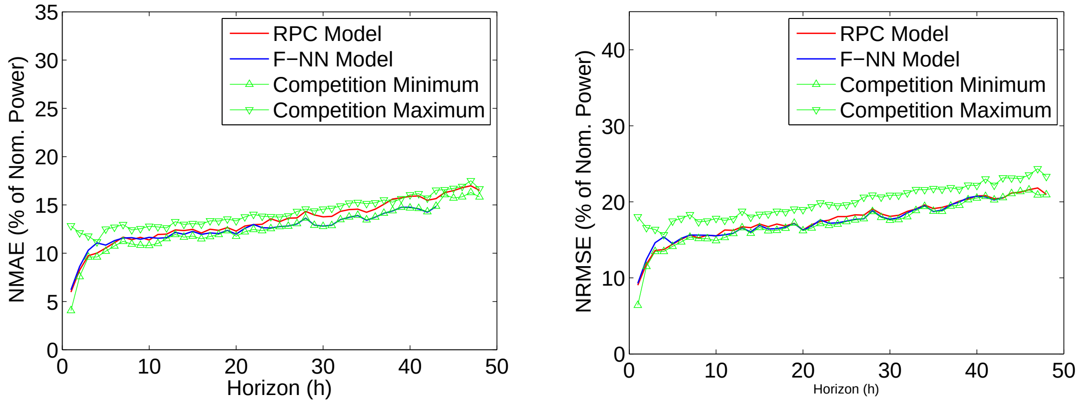
The two modelling approaches are illustrated in Figure 4.5. Approach (*a*) is a cluster approach, where the wind speed forecasts of the reference wind farms in each cluster are averaged and then used as input to a single RPC model. No production measurements are considered as input. The cluster forecasts obtained for each cluster are then summed to obtain the final regional forecast. Approach (*b*) is a cluster approach where the wind speed forecasts provided for each wind farm are used as input to an RPC model which computes a forecast of the aggregated production of the wind farms in the cluster. The cluster forecasts computed by each RPC model are then averaged to obtain the final cluster forecast. As with approach (*a*) the cluster forecasts are summed to obtain the regional production forecasts.

4.6 Evaluation of Regional Model Performance

4.6.1 RPC Model Validation

In this section the performance of the RPC model is compared to that of other state-of-the-art models. This comparison is performed on three single wind farm forecasting case studies used in the Anemos Competition. The case-studies evaluated here are: Tunø Knob off-shore wind farm, Golagh wind farm located in complex terrain, and Alaiz wind farm located in very complex terrain.

Given the observed performance of the RPC model we limit the detailed examination of



(a) Tunø Knob NMAE

(b) Tunø Knob NRMSE

FIGURE 4.6: *NMAE and NRMSE as a function of forecast horizon of the RPC and F-NN model for Tunø Knob wind farm. The green curves represent the minimum and maximum values obtained by the state-of-the-art models, which participated in the Anemos Competition.*

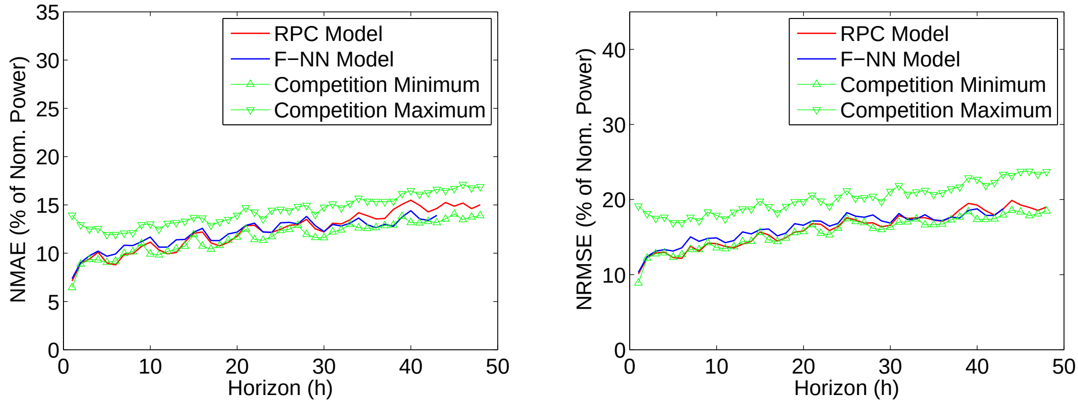
results to the NMAE and NRMSE criteria for concision. The differences observed between the RPC model and the F-NN model for other criteria are comparable to those found for the NMAE and NRMSE. The results obtained for other criteria can be found in Appendix B.1.

Tunø Knob

In Figure 4.6 the normalized mean absolute error (NMAE) and the normalized root mean square error (NRMSE) as functions of the forecast horizon are presented for the case of Tunø Knob wind farm. The green curves with triangular markers represent the minimum and maximum values of the error criteria obtained by the models participating in the Anemos competition. These curves do not necessarily reflect the performance of a single model but rather are the best and worst performances observed for a given horizon on the ensemble of state-of-the-art models.

From these results it is clear that all the models have very similar performance on this case. Indeed, the minimum values of NMAE range from 4% to 15.8% of nominal power, while the maximal values range from 12.8% to 16.66%. Apart from the first forecast horizons, the envelope of performance is only 1 to 1.5 percentage points wide. The same behaviour can be observed for the NRMSE, where for most horizons the performance envelope is 1.5 to 2 percentage points wide. The superior width observed for the first horizons can be explained by the fact that some models in the competition do not fully exploit on-line power measurements, while others such as the F-NN and RPC model exploit this information. Hence, some models are penalized for the first horizons, while for higher horizons the performance is more homogenous due to the use of NWP forecasts, which all models try to exploit to the fullest.

When examining the performance of the F-NN and RPC models more closely, it can be noticed that both models have very similar performance. In terms of NMAE and NRMSE the



(a) Golagh NMAE

(b) Golagh NRMSE

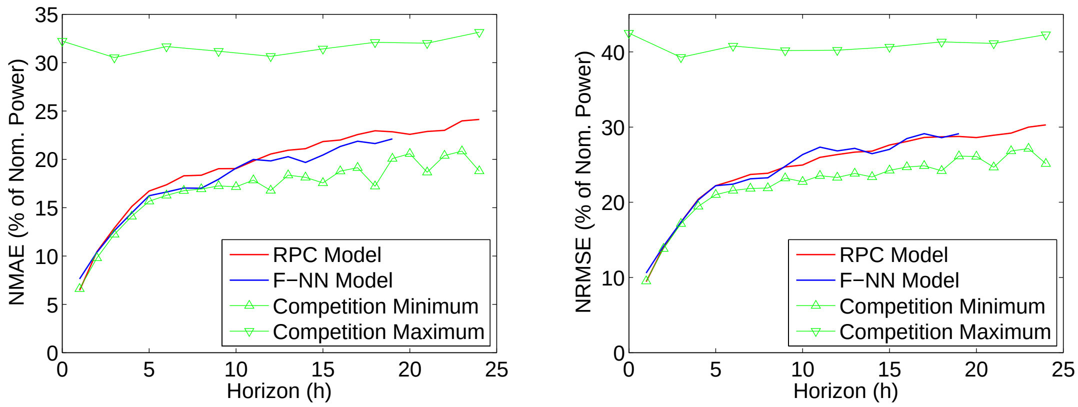
FIGURE 4.7: *NMAE and NRMSE as a function of forecast horizon of the RPC and F-NN model for Golagh wind farm. The green curves represent the minimum and maximum values obtained by the state-of-the-art models, which participated in the Anemos Competition.*

RPC model slightly outperforms the F-NN model for the first six horizons. After the sixth horizon both models have almost identical performance up to horizon 12. From horizon 12 onward, the performance of the F-NN model outperforms the RPC model in terms of NMAE. This behaviour is not as evident in terms of RMSE. When examining the RMSE, we can notice that the F-NN model sometimes clearly outperforms the RPC model for some horizons, while for other horizons the performance difference is hardly noticeable.

Golagh

In Figure 4.7 the NMAE and the NRMSE as functions of the forecast horizon are presented for the case of Golagh wind farm. In this case all the models also have similar performance, with performance envelopes 2 to 3 percentage points wide for the NMAE and 4 to 5 percentage points wide for the NRMSE for most forecast horizons. These results clearly present a higher spread than those for Tunø Knob wind farm, yet the general shape of the minimum and maximum curves is similar in both cases. The higher spread of model performance can be explained by the complexity of the terrain. In [32] it was found that model spread is highly influenced by terrain complexity; model spread is proportional to terrain complexity. Tunø Knob wind farm being an offshore wind farm, its “terrain” can be considered as flat, (although with varying roughness length, whereas Golagh is located in a hilly area in Ireland. However, the minimum error criteria is lower for Golagh than for Tunø Knob, this might be linked to more pronounced wind park effects and to varying roughness length often observed in off-shore wind farms.

With respect to the performance of the F-NN and RPC models, we can note that their performance for both NMAE and NRMSE are very close and sometimes equivalent to the minimum obtained in the competition. When examining the performance of the RPC model, we can notice the same under-performance in terms of NMAE with respect to the F-NN



(a) Alaiz NMAE

(b) Alaiz NRMSE

FIGURE 4.8: *NMAE and NRMSE as a function of forecast horizon of the RPC and F-NN model for Alaiz wind farm. The green curves represent the minimum and maximum values obtained by the state-of-the-art models, which participated in the Anemos Competition.*

model observed in the previous case, for higher horizons. In the present case, the difference becomes noticeable for horizons above 30 hours ahead. For shorter horizons, the RPC model presents slightly better performance than the F-NN model. In terms of NRMSE, the same behaviour can be observed, with the RPC model slightly outperforming the F-NN model for horizons below 30 hours and presenting higher errors for horizons above 30 hours.

Alaiz

In Figure 4.8 the NMAE and the NRMSE as functions of the forecast horizon are presented for the case of Alaiz wind farm. As mentioned in the preamble of this section, the minimum and maximum values of NMAE and NRMSE clearly illustrate the difficulty of predicting for this wind farm. The NMAE minimum values range from 6.6% to 18.8% of nominal power for horizons ranging from 1 to 24 hours ahead. For the first horizons these values are similar to those found for the other cases. However, the performance rapidly deteriorates, with values for the 24-hour horizon higher than those found for the 48-hour horizon in the other cases. Further, the spread of model performance is very important, the width of the performance envelope varies between 30 to 15 percentage points. Clearly, some models have a very difficult time predicting for this farm. This behaviour can also be observed for the NRMSE criterion.

The performance of the F-NN and RPC models us very similar for both criteria. In terms of NMAE, the F-NN presents a relative advantage for horizons between 6 and 19 hours ahead. However, when examining the RMSE, the difference between both models is not as clear-cut. The RPC model presents a rather smooth increase of the RMSE as the horizon increases. The performance of the F-NN model is more variable, it seems to oscillate around that of the RPC model; for some horizons the RPC model presents lower NRMSE, while fore others the F-NN has better performance.

The overall performance of the F-NN and RPC models compares favourably with the maximum and minimum error values for both NMAE and NRMSE. For both models, the NMAE and NRMSE values are below the middle of the competition performance envelopes. It should be noted however that the maximum error values for both criteria in this case are very high and, for higher horizons, comparable to those that can be obtained with the Persistence model.

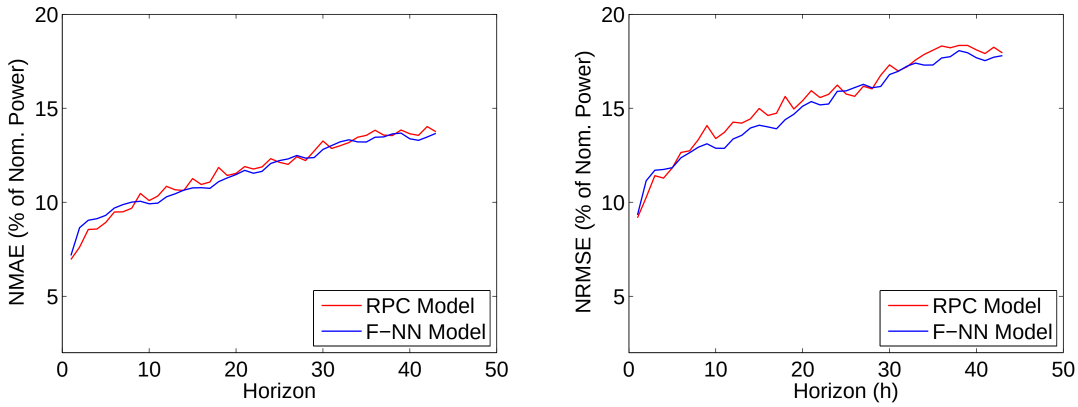
When comparing the F-NN and RPC model performance to the minimum values we can note that for the first horizons these models have near-minimum values for both criteria. For horizons higher than 6 hours, the performance of both models decreases. This can be explained by the fact that the explanatory data is not fully exploited by these models. Indeed, in the case of Alaiz, wind direction is an important explanatory variable. By definition, the RPC model does not exploit this variable, hence it is penalized when compared to other models. On the other hand, the F-NN model does exploit this variable, but defines a transfer function that does not take into account the forecast horizon. In the present case both the wind direction and horizon information are very important; other models that explicitly consider the forecast horizon and the wind direction present better performance.

Conclusions

From the results presented above, some conclusions can be formulated regarding the performance of the RPC model. From a general perspective, the performance of this model compares favourably to that of other state-of-the-art models. The RPC model has generally good performance, in most cases its performance is very close to that of models with more advanced functionalities such as on-line adaptation. Generally speaking, the RPC model's performance is in line with that of other commercially available wind power forecasting models.

An important aspect is the similarity of behaviour of the RPC model and that of other models tested in the Anemos competition. Both the NMAE and NRMSE values computed for this model follow the same trend as that observed for the other models. The values of different moments of the model error distribution (bias, standard deviation of the error) of the RPC model and those published in [36] for the cases of Tunø Knob and Golagh, are very similar. When most models perform well and the performance envelope is small, the RPC models performance is well within the envelope. When the behaviour of the competition models is more variable and the performance envelopes are wider, the RPC model's performance is within the envelope and tends to be situated near the best possible performance.

When specifically comparing the RPC and the F-NN, it is difficult to determine a definite advantage of one model over the other. When the performance of all models is very similar, the differences between both models are very tenuous. When the performance is more variable, as in the Alaiz case, no definite difference can be established. Indeed, for some performance evaluation criteria (NMAE) the F-NN model presents better results for some horizons, while for other evaluation criteria (NRMSE) the RPC model performs bet-



(a) NMAE (b) NRMSE

FIGURE 4.9: *NMAE and NRMSE of the RPC and F-NN models for the direct upscaling approach for the Irish case. The input data used by both models consists of the NWP forecasts and on-line power measurements for wind farm 8.*

ter for some horizons and the F-NN model for others. In a general sense, the RPC model benefits from its horizon dependent structure, while the F-NN benefits from its capacity to incorporate more explanatory variables in its forecast computation process.

From the above we can conclude that the RPC model is generally comparable to other models in the literature both in terms of performance and presents the behaviour as other models depending on the case study. Thus, we can reasonably base part of the studies conducted in this thesis on the use of the RPC model. In particular, the similarity in the performance of the RPC and F-NN models, and the difficulty in determining which, if any, is superior, leads us to prefer the use of the RPC model in many cases. Indeed, this model has the advantage of requiring very short off-line computation time when compared to that of the F-NN model. In the frame of the research carried in this thesis, having a computationally light model offers the advantage of allowing the examination of many different cases. This would not have been possible with a more computationally demanding model such as the F-NN model.

4.6.2 Evaluation of the Direct Regional Forecasting Approach

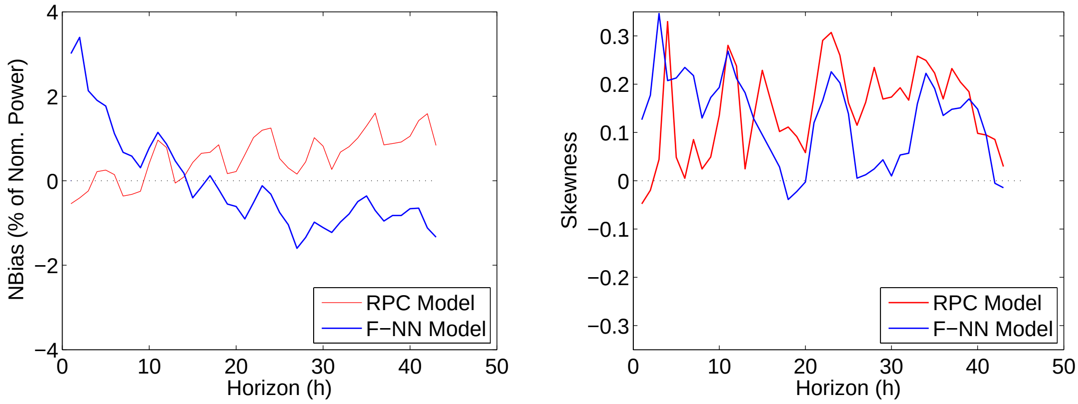
As described in subsection 4.5.2, two cases were examined to study the properties of the direct upscaling approach. Given the similarity of the results between both cases only the results obtained for the Irish case are described here. The results for the Danish case can be found in Appendix B.2.

In Figure 4.9, the NMAE and NRMSE of the RPC and F-NN models as a function of the forecast horizon are presented for the direct upscaling approach for the Irish case. These results were obtained by training the F-NN and RPC models on the NWP and on-line production measurements available for wind farm 8, a.k.a. Golagh wind farm. Hence, the models are extrapolating the regional production from data provided for a single wind farm.

The first aspect that can be noted is that the NMAE and NRMSE values for this case are relatively lower than those observed in the single wind farm cases examined in the previous section. This can mainly be explained by the smoothing effect which renders forecasting regional production "easier" than forecasting single wind farm production. Further, the general shape of the NMAE and NRMSE curves is the same as that found for single wind farm forecasts. For low horizons, 1 to 6 hours ahead, the error increases sharply and then increases more regularly as the forecast horizon increases. As with the single wind farm case, this can be explained by the relative influence of the on-line production data and the NWP data on forecast performance. For low horizons, the forecasts are heavily influenced by the past production measurements, whereas for higher horizons, the forecasts are mainly based on the transformation of NWP data to power. As can be seen in the correlograms presented in subsection 3.6.1, the auto-correlation of the production rapidly decreases for the first few time lags, this behaviour is translated into the short-term increase in the NMAE and NRMSE values, with the added effect of estimating the regional production from the production measured at a single wind farm. For higher horizons, the steady increase of the error can be linked to reduction in mutual information between the NWP data and the regional production observed in Figure 3.15. Further, we can notice that the mutual information between wind speed and regional production "peaks" for the 6-hour ahead horizon. For this same horizon the autocorrelation falls below 0.8, which might explain the fact that it is roughly after this horizon that the behaviour of the NMAE and NRMSE curves changes.

When comparing the performance of the RPC and F-NN models, we can note that both models present the same behaviour for both the NMAE and NRMSE. For the first horizons the RPC model performs better. This changes after horizon 10 for the NMAE and horizon 5 for the NRMSE, and continues until horizon 25 for both criteria. After this horizon, the relative performance of the models alternates on a much shorter period. This behaviour can be in part explained by the fact that the RPC model better exploits the on-line power measurements for the first horizons due to its ability to take into account the forecast horizon. However, for higher horizons, the ability of the F-NN model to exploit other input variables than just the wind speed gives it a slight advantage. This derives in part from the fact that wind direction at the wind farm provides complementary information on the prevailing meteorological situation over Ireland, which can be exploited by the F-NN model. However, for horizons superior to 25 hours ahead it is likely that the error in the NWP forecasts outweighs the wind direction advantage of the F-NN model, which leads to the observed variability in the performance of the models.

In Figure 4.10 the normalized bias and the skewness of the model errors are plotted versus the forecast horizon. Here we remind the reader that the bias is the average of the model errors, or the first moment of the error distribution and that the skewness is the third moment of the error distribution. The bias indicates if, for a given horizon, the models systematically under- or over-estimate the power production. If the bias is positive, then the model is under-estimating the future production and conversely a negative bias indicates an over-estimation of the production. The skewness indicates the degree of symmetry of the error distribution. If the skewness is null then the distribution is symmetrical, if the



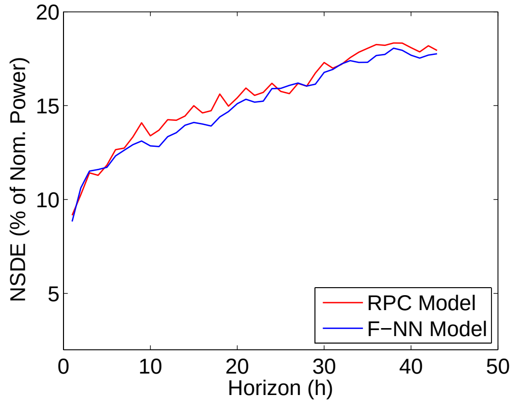
(a) Normalized Bias (b) Skewness

FIGURE 4.10: *Normalized Bias and Skewness of the RPC and F-NN models for the direct upscaling approach for the Irish case.*

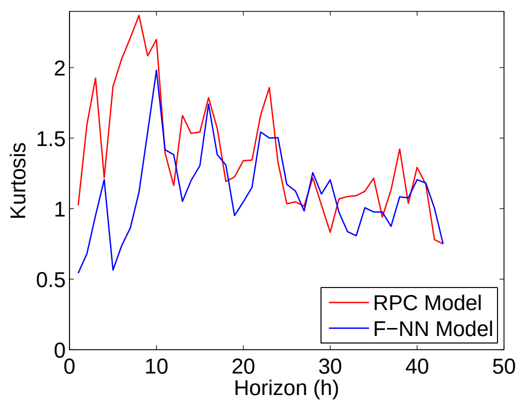
skewness is negative then the distribution is left-skewed (the left tail of the distribution is the "longest") and if the skewness is positive the distribution is right-skewed (the right tail of the distribution is the "longest").

Concerning the bias, it can be noted that the evolution of the bias with the forecast horizon significantly differs for both models. The F-NN model has an important positive bias for the first horizons, which then decreases and is near zero around horizon 12. The F-NN's bias then becomes increasingly negative. The evolution of the RPC's bias is almost opposite to that of the F-NN. The RPC's bias is slightly negative for the first horizons and increases steadily as the forecast horizon increases. The RPC's bias is also near zero for horizons around horizon 12, after horizon 12 the bias becomes positive. The bias of the models evolves in opposite directions, one steadily increases while the other presents a rather sharp initial decrease and then steadily decreases. This difference can possibly be explained by the fact that the RPC model considers the forecast horizon as an explanatory variable while the F-NN model is a recursive model that does not consider forecast horizon as an explanatory variable. When the meteorological error becomes predominant, for horizons superior to 12 hours ahead, we can notice that the biases of the models are positively correlated. Moreover, the absolute value of the bias is similar for both models and lower than that typically observed for single wind farm forecasts.

The skewness values of both models are very similar, and present predominantly positive skewness. In general, single wind farm wind power prediction models present positive skewness. The fact that skewness is generally positive for single wind farm predictions can be linked to the nonlinear and bounded shape of the power curve as well as the fact that wind speeds are predominantly low [36]. Evidently, the same phenomenon takes place in the frame of regional wind power forecasting. The difference in behaviour between the two models closely follows that observed for the NMAE and the NRMS curves with F-NN having higher skewness for the first horizons and RPC having slightly higher skewness for higher



(a) NSDE



(b) Kurtosis

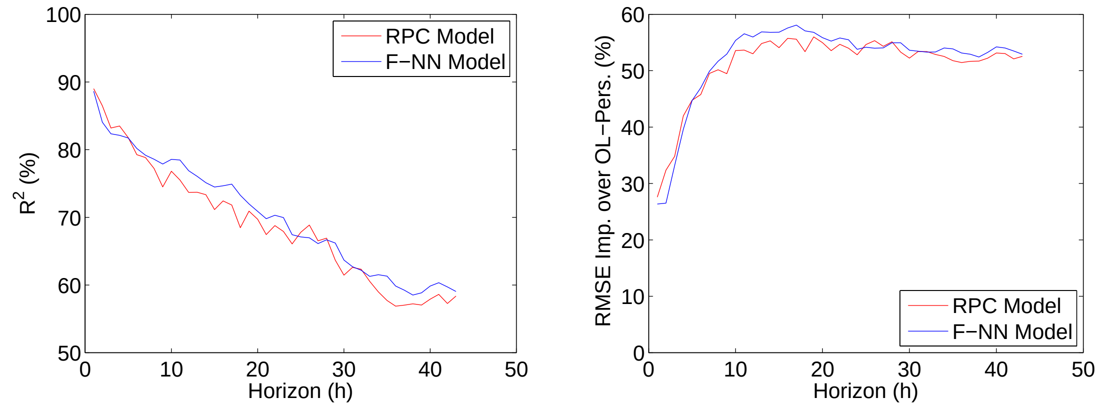
FIGURE 4.11: *NSDE and Kurtosis of the RPC and F-NN models for the direct upscaling approach for the Irish case.*

horizons. The skewness values are relatively low and comparable to those found for forecast error distributions in single wind farm forecasting cases.

In Figure 4.11 the normalized standard deviation of the errors (NSDE) and the kurtosis of the error distributions are plotted as functions of the forecast horizon. These two measures constitute the second and fourth moments of the error distributions. The NSDE provides information on the "width" of the error distribution, which can be considered as quantifying to a certain extent the uncertainty of the prediction. The kurtosis provides information on the shape of the error distribution with respect to a Gaussian distribution. If the kurtosis is positive, the distribution presents a sharper peak around the mode and longer tails than the Gaussian distribution. Conversely, if the kurtosis is negative the distribution presents a flatter peak around the mode and shorter tails.

We can notice that the general evolution of the NSDE and NRMSE values for both models are very similar. This is due to the fact that the NSDE is an NRMSE computed on the centred error series. The behaviour that can be observed from the NSDE is very similar to that observed through the study of the NRMSE. As the forecast horizon increases the NSDE and by extension the prediction uncertainty increases. The shape of the NSDE — with low initial values rapidly increasing for the first horizons and then steadily increasing for higher horizons — is directly influenced by the on-line power measurements and NWP forecasts on the prediction error for the different horizons. Generally speaking both the RPC and F-NN model have comparable performance in terms of SDE, the difference in performance are linked to the model's properties as described for the NRMSE.

The evolution of the kurtosis is only moderately influence by the prediction horizon. In the present case we can notice that the kurtosis values seem to initially increase up to horizons 6 to 12 and then seemingly decrease as the horizon increases. For the first horizons, the models make slightly more low errors than high errors. For higher horizons, the models commit many small errors, but few very large errors, this explains the observed sharpness



(a) R^2

(b) Improvement over OL-Persistence

FIGURE 4.12: *R^2 and NRMSE improvement over OL-Persistence of the RPC and F-NN models for the direct upscaling approach for the Irish case.*

of the error distribution as measured by the kurtosis, as well as the higher NSDE since this criterion strongly penalizes larger than average errors. For higher horizons, the kurtosis values are rather constant, the shape of the error distributions varies only slightly toward that of a Gaussian. In the present case we can note that forecast horizon seems to have a stronger influence on kurtosis than in the single wind farm forecasting case [36]. Further, we can notice that both models present kurtosis values in the same value ranges, although the variations with respect to forecast horizon do not systematically present positive correlation.

In Figure 4.12 the R^2 skill score and the improvement over OL-Persistence (defined in autorefeq:onlinepers) in terms of NRMSE are presented for the F-NN and RPC models as a functions of the forecast horizon. The R^2 score measures the amount of variance of the predicted variable explained by the forecast model. Ideally this score should be equal to 100% for all horizons. The improvement over OL-Persistence indicates the benefit derived from the use of an advanced model over a simple model in terms of forecast error reduction. If the advanced model provides perfect forecasts, that is forecasts with null error, 100% improvement is achieved.

We can notice that, for both models, the R^2 criterion decreases as the forecast horizon increases. The shape of the R^2 curves is similar to that observed for single wind farm forecasting cases, and reflects the increasing level of forecast error observed as forecast horizons increase. When comparing the RPC and F-NN results, these are once again very similar. The relative level of skill of each model is comparable to that observed for other criteria such as the NRMSE. For the first few horizons, the RPC model holds the advantage, then the F-NN presents higher performance, and after horizon 25 the difference becomes less pronounced.

With regard to the improvement, in terms of NRMSE, over OL-Persistence, we can notice that for small horizons, improvement rather low, but higher than that typically observed when comparing advanced models to Persistence in single wind farm forecasting cases. This can be explained by the fact that in the OL-Persistence model, the ratio of the reference farm's nominal power on the total regional capacity does not constitute a good extrapolation factor. However, as the forecasting horizons increase, the improvement of both models over OL-Persistence becomes more important. It must also be noted that the improvement over OL-Persistence reaches a maximum plateau for horizons between 12 and 24 hours ahead. After horizon 24 the improvement decreases somewhat. This is the translation of the NWP model accuracy, which usually decreases for horizons superior to 12-24 hours ahead. From a general perspective, it is clear that using an advanced model clearly improves forecast accuracy, even for short forecast horizons. When comparing the performance of the RPC and F-NN model, it is clear that the improvement over OL-Persistence measured for each model reflects the RMSE performance of both models. The two curves present the same characteristics as the RMSE curves described earlier.

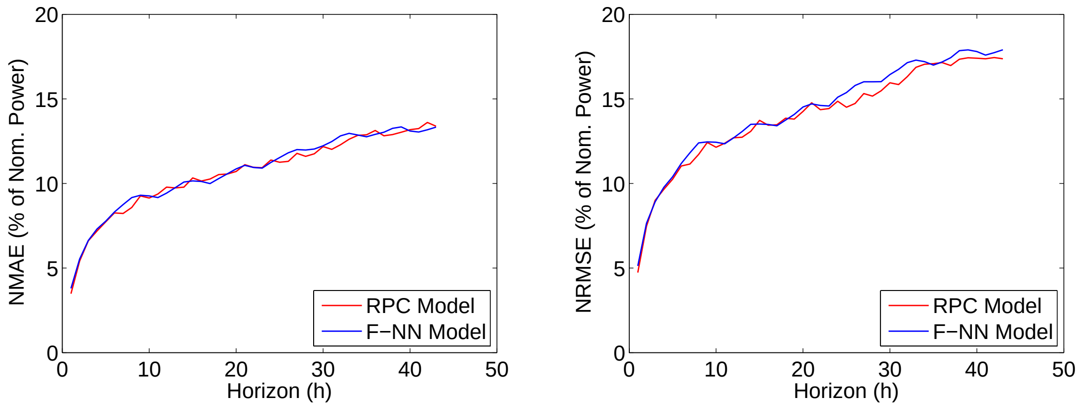
Conclusions

The first conclusion that can be reached from the results of the direct forecasting scheme is that the F-NN and RPC models are both capable of predicting regional production in a direct fashion. Indeed, the results, in terms of forecast performance, are better than those obtained for single wind farm forecasting in the literature. This good performance can be mainly attributed to the smoothing effect present in regional forecasting which facilitates the task by reducing the variability of the dependent variable.

A second conclusion that can be reached is that the difference in forecast performance between both models is not very important. In the Irish case, where only one reference wind farm provided data to the models, a slight advantage can be accorded to the F-NN model given that it has the best performance for most horizons. However, in the Danish case, the RPC holds the advantage. This behaviour is consistent with that observed in the Anemos competition where no forecasting model clearly outperformed the others both in terms of terrain type and forecast horizon. This behaviour has also been observed in other more general forecasting competitions such as the M-3 competition where no overwhelming advantage can be detected between advanced models from one case to another [115, 116].

From a more specific perspective the Irish cases examined here shows that it is possible to forecast regional wind power production from data provided by a single reference wind farm with reasonable accuracy. The performance results found for this case are in line and even slightly better than those found for the Golagh case presented in subsection 4.6.1. It should be noted however, that these results were obtained through the use of data from a reference farm whose choice was not random. Indeed, Golagh wind farm has a nominal power of 15 MW out of a total of 70 MW for the region and is located relatively close to other farms that constitute the majority of the installed capacity in the case.

From the Danish case it can be concluded that using averaged NWP variables as explanatory variables does not lead to low performance. The results obtained in the case of direct regional forecasting from averaged NWP data in the Danish case lead to results that compare very favourably to those obtained for single wind farm forecasting. This can be



(a) NMAE (b) NRMSE

FIGURE 4.13: *NMAE and NRMSE of the RPC and F-NN models for the cascaded approach for the Irish case.*

mainly attributed to the very important smoothing resulting from the large number of wind farms installed in the Jutland Funen area. Another important factor that can explain these results is the homogeneous nature of meteorological situations in Denmark. The average wind speed likely reflects the wind speeds witnessed by all wind turbines in the region.

4.6.3 Evaluation of the Cascaded Regional Forecasting

As described in subsection 4.5.2, the Irish case was examined to study the properties of the cascaded regional forecasting approach. In this case all the wind farms in the region were considered as reference wind farms for which NWP data and production measurements are available. In this approach the RPC and F-NN models were used to generate single wind farm forecasts for all the wind farms in the region. The single wind farms forecast were then summed to produce the regional wind power forecasts. This approach differs from the direct approach by the number of parameters estimated in the model and by the number of input variables available to the model. The fact that the model possesses many parameters is directly linked to its structure. Indeed, for each wind farm, a single F-NN or RPC model was tuned. This, when considered as a whole, translates to a very important number of parameters. The fact that the regional forecasting model, as a whole, relies on significantly more input variables, results in an increase of the number of model parameters and, most importantly, permits an important reduction of the forecasting error.

In Figure 4.13 the NMAE and the NRMSE obtained using the RPC and F-NN models in the cascaded approach are presented. We can notice that for both criteria the shape of the curves is very similar to that found for the direct approach and more generally for the single wind farm forecasting approaches. Further, both models present NMAE and NRMSE values which are very close. The values only slightly differ for horizons 25 to 35. When compared to the values obtained using the direct approach we can notice that the NMAE and NRMSE found for the cascaded approach are noticeably lower.

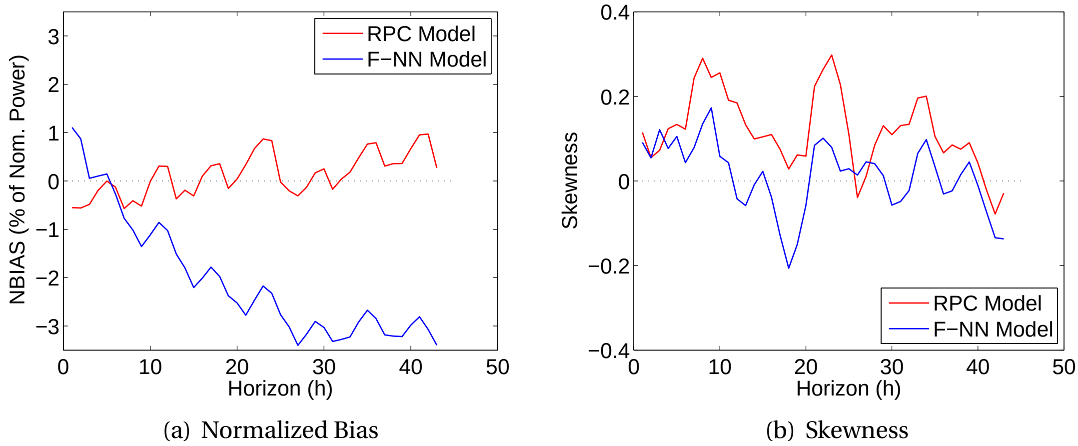


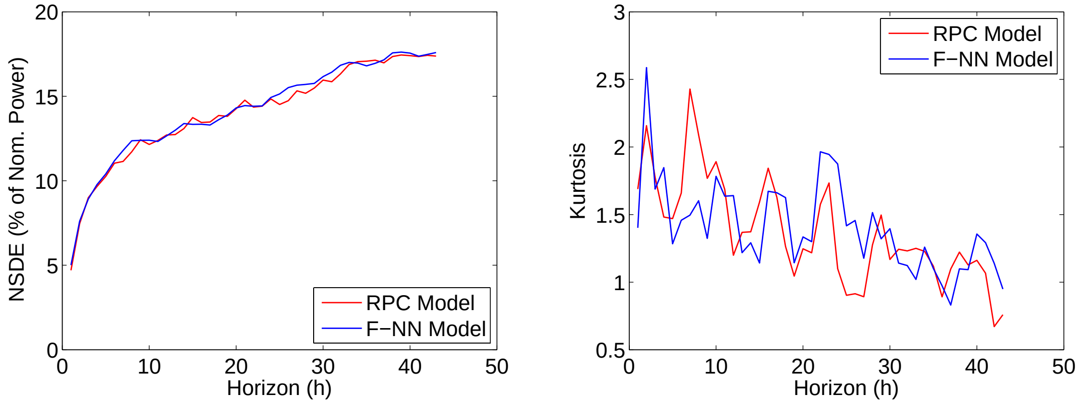
FIGURE 4.14: *Normalized Bias and Skewness of the RPC and F-NN models for the cascaded approach for the Irish case.*

It can be noted that the NMAE presents a sharp gradient for the first forecast horizons. For horizons 1 to 10 the NMAE values increase very rapidly when compared to the initial increase observed for the direct approach. This can be linked to the availability of on-line production measurements for all wind farms considered for this case. This translates to much lower absolute error values for the first horizons: less than 5% in the present case compared to more than 6% in the direct case for the 1 hour ahead horizon. Both models adequately exploit the measured data to lower the forecast error in the first horizons. With respect to higher horizons we can notice that the difference with the direct approach is less pronounced. It seems that having NWP forecasts for several locations in the area does not provide the same improvement for the long-term horizons as does the availability of on-line production measurements for the short-term horizons.

The observations formulated for the NMAE also apply to the NRMSE. The reduction of the error when compared to the direct approach is more pronounced for the short-term horizons than for the long-term ones. However, for the first horizons, the reduction in NRMSE is more important than that observed for the NMAE. For the first horizon the NRMSE drops from 9% in the direct case to 5% in the cascaded case. The reduction is roughly of 4 percentage points for the NRMSE and only of 1 percentage point for the NMAE. This indicates that having complete on-line production measurements as input to the models reduces the average error, but more importantly, it reduces the number of very large errors.

In Figure 4.14 the normalized bias of the forecast errors and the skewness of the error distributions are presented for the cascaded approach. For both criteria the values are of the same order as those observed in the direct approach. However, a marked difference can be observed for the F-NN model.

The bias values observed for both models follow the same general trend as those observed in the direct approach. The RPC model bias slightly increases with forecast horizon.



(a) NSDE

(b) Kurtosis

FIGURE 4.15: *NSDE and Kurtosis of the RPC and F-NN models for the cascaded approach for the Irish case.*

In the case of the F-NN model the bias is positive for the first horizons and decreases with forecast horizon. The main difference with the results from the direct approach are a much more pronounced decrease in bias. In the direct approach the bias observed for the last horizon for the F-NN model was equal to -1.8% of nominal power, in the present case the bias for the last horizon is almost -3.5% of nominal power. This increase of the bias can be linked to the additive nature of the cascaded approach. For all wind farms in the region, the single wind farm forecasts computed with the F-NN model present bias values with the same trend. Given that all models over-predict the production of the single wind farms, when the single forecasts are summed to obtain the regional prediction, the production is necessarily over predicted. This behaviour is not observed for the RPC model because the error biases obtained for the single wind farm forecasts present differing trends. Therefore, when the single wind farm RPC forecasts are summed, the systematic error is smoothed.

With respect to the skewness of the error, a difference in behaviour between the F-NN and RPC model can also be noticed. The skewness of the RPC model errors is mostly positive. This comparable to the skewness observed for both model in the direct forecasting approach. In the case of the F-NN model, the skewness values oscillate around zero. This can once again be linked to differences in the single wind farm forecast errors. The RPC model tends to produce errors that are more consistently positively skewed, than those obtained using the F-NN model. It can be noted however that the difference between both models, in absolute terms, is low.

In Figure 4.15 the normalized standard deviation and the kurtosis of the error distributions as a function of the horizon are presented. For both criteria we can notice that the shape of the curves differs from that found in the direct case. The NSDE presents lower values and the kurtosis presents a marked downward trend. Here the shape and values observed for the NSDE have the same explanation than that for the NRMSE, we will therefore more specifically concentrate on analysing the kurtosis.

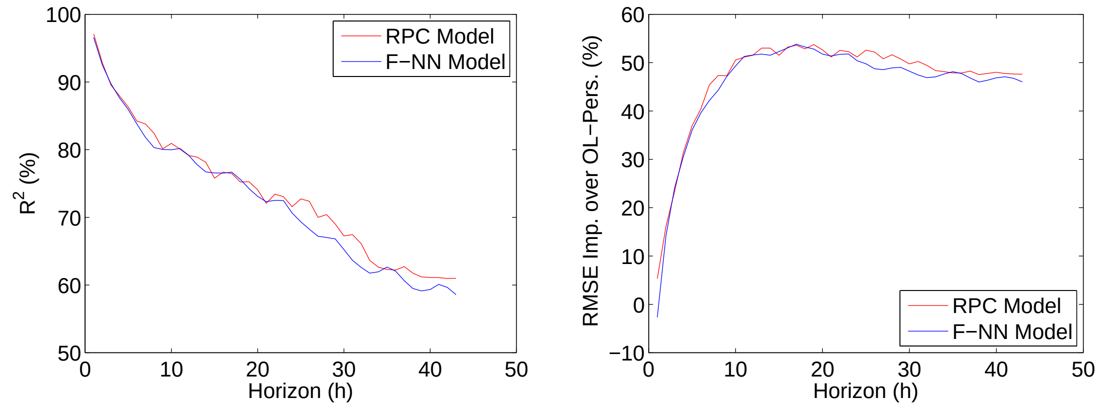
In the present case the kurtosis presents a clear horizon dependent downward trend. The error distributions are sharp for the first horizons and progressively evolve toward normally shaped distributions. This behaviour goes against the findings presented by Pinson in [36] for single wind farm forecasts. Indeed Pinson found that the forecast horizon had but negligible impact on the kurtosis of the error distributions in the single wind farm case. This behaviour can possibly be explained by a stronger autoregressive component of regional production with respect to single wind farm production. Indeed, Pinson showed that, in single wind farm forecasting, high initial kurtosis for the two first horizons was due to the use of power measurements as input to the models [36]. In the present case the regional models use as input the measurement of the total regional production. Therefore, high kurtosis values for the short term are to be expected. However, given the strong autoregressive nature of regional production (see subsection 3.6.1), the effects of the power measurements are present, to a progressively lesser degree, for higher horizons than in the single wind farm forecasting case. Another explanation can be that in the present test-case, this behaviour is present for some, or even most, single wind farm forecasts used to compute the total. Given the dependence which exists between forecast errors obtained for spatially distributed wind farms [24], the summation process will probably not lead to normally distributed errors as could be expected from the central limit theorem.

In Figure 4.16 the horizon dependent R^2 and improvement over OL-Persistence are presented for the cascaded case. OL-Persistence being a generalisation of Persistence, when complete on-line production measurements are available both naïve models are equivalent. Therefore, in the present case the improvement over OL-Persistence is equivalent to the improvement over Persistence. The performance difference for both criteria observed between the F-NN and RPC model is small and comparable to that already observed for the other evaluation criteria presented above. It can be noticed that the values obtained for both criteria are significantly different from those obtained in the direct forecasting approach.

The R^2 values obtained in the cascaded case present the same general evolution as those observed in the direct approach. The main difference with the previously described approach is the values observed. In the direct case the R^2 values decreased from 90 % for the first horizon to roughly 58 % for the last prediction horizon. In the present case the values decrease from 97 % down to 60 %. There is clearly an improvement over all forecasting horizons. The improvement is more important for the first horizons where complete online production measurements clearly improve forecast quality. For higher horizons, where on-line information is less critical to forecast performance, the added information only provides slight improvement.

The improvement over OL-Persistence for the cascaded approach presents a marked difference with the one found for the direct approach. The values for the first horizons are much lower than the ones found in the direct approach: 0 to 5% for the cascaded approach versus 25% for the direct approach. For higher horizons the values obtained in the present case are also lower than those observed in the direct approach.

These lower values can be explained by the performance improvement of OL-Persistence



(a) R^2 (b) Improvement over OL-Persistence

FIGURE 4.16: *R^2 and NRMSE improvement over OL-Persistence of the RPC and F-NN models for the cascaded approach for the Irish case.*

over all horizons. For the short-term horizons, the low improvement values are due to the fact that the OL-Persistence model presents very high accuracy. This high accuracy can be linked to the production smoothing effect, which translates to very high autocorrelation values for lags below 6 hours (see subsection 3.6.1, Figure 3.6). The high autocorrelation leads to good performance for autoregressive models such as the OL-Persistence. In the case of direct forecasting presented previously, the “advanced” models are capable of beating OL-Persistence in part due to the meteorological information and to a better estimation of the factor linking the wind farm’s production and the regional production. When complete production measurements are available, as in the present case, the advantage of the more sophisticated models over OL-Persistence is less evident, and the improvement values observed for the first horizons are comparable to those typically found in for the improvement over Persistence in single wind farm forecasting cases.

Conclusions

The first conclusion that we can draw from the results presented for the cascaded regional forecasting approach is that this approach leads to globally good quality forecasts. The cascaded significantly outperforms the results presented for the direct approach. The improvement of forecast performance in the present case can be linked to the larger number of explanatory variables used to forecast the regional production. The cascaded model benefits from more detailed information on the production at the time of forecast computation and also more detailed information on the likely evolution of the atmosphere over the region.

The most important improvement in terms of performance takes place for short-term horizons. This primarily results from the more precise knowledge of the regional production at the time of forecast computation. Indeed as shown in subsection 3.6.3 having on-line production measurements from only a few reference wind farms can allow to quite precisely estimate the current production and hence the total production of a region. The superior amount of on-line information also leads to better OL-Persistence forecasts. Because of this, the relative gain of the cascaded approach with respect to OL-Persistence is lower. However it must be noted that OL-Persistence can be beaten even for the first horizons. Indeed, statistical models are routinely used to evaluate the present wind power injected in networks faced with large numbers of wind turbines [86].

The performance increase resulting from the increased number of NWP variables used in the present model is less important than that observed from the increase of on-line production data. The difference in performance increase, observed between the short and long term horizons, is obviously due to the fact that for long-term horizons, the main information exploited by the forecasting models are the NWP forecasts. As shown in [24], the correlation of errors in the single wind farm power forecasts increases with forecast horizon, this automatically leads to a reduction of the error smoothing due to the increase of the covariance term of the forecast error variance (see section 3.3). From this, it can be conjectured that by adequately selecting the forecasts used to compute the higher horizon forecasts, some improvement in accuracy can be achieved. Indeed, by selecting single wind farm forecasts presenting low error correlation, a more important error smoothing can possibly be achieved, which might lead to better overall performance.

The additive nature of the cascaded function used in the present case compounds the error properties of the single wind farm forecasts used as input. This can be noticed in particular for the normalized error bias results for the F-NN model where consistently negative biases in the single wind farm forecasts lead to a negative bias of the regional forecast. In the case of the RPC model, the single wind farm forecasts do not present such a homogeneous behaviour, hence, the sum of the forecasts presents a bias that is centred around zero for all horizons. From this, it must be concluded that, when using this type of approach, care should be taken to adequately characterize the single wind farm forecasts in order to foresee possibly undesirable behaviour resulting from forecast aggregation.

Finally, we can conclude that the difference in forecast performance between the F-NN and RPC models is again very slight in the regional forecasting approach. The difference in performance observed is comparable to that observed in the direct approach and the single wind farm forecasting cases. Neither model seems to possess a clear advantage over the other.

4.6.4 Evaluation of the Cluster Regional Forecasting

As described in subsection 4.5.2, the cluster approach is evaluated by dividing the Irish wind farms into three groups, or clusters. For each group a F-NN and a RPC model were trained to provide forecasts of the aggregated production of the wind farms in the group. The sub-regional cluster forecasts were then summed to obtain the total regional production. The NWP data of all the wind farms was considered as input to the models. Only wind farm 8 (North Cluster) is considered as providing on-line production measurements, which were only used as input to the models tuned to provide forecasts for the North Cluster. This approach differs from the cascaded approach in that it considers less on-line production data, see Table 4.1. The approach differs from the direct approach by the number of NWP

TABLE 4.1: *Input data considered in the different regional forecasting approaches tested on the Irish case study*

Input data	Approach		
	Direct	Cascaded	Cluster
NWP data	WF 8	All	All
Production data	WF 8	All	WF 8 (Only for North Cluster)

variables considered and by the fact that the dependent variable (total regional production) is decomposed into three aggregates.

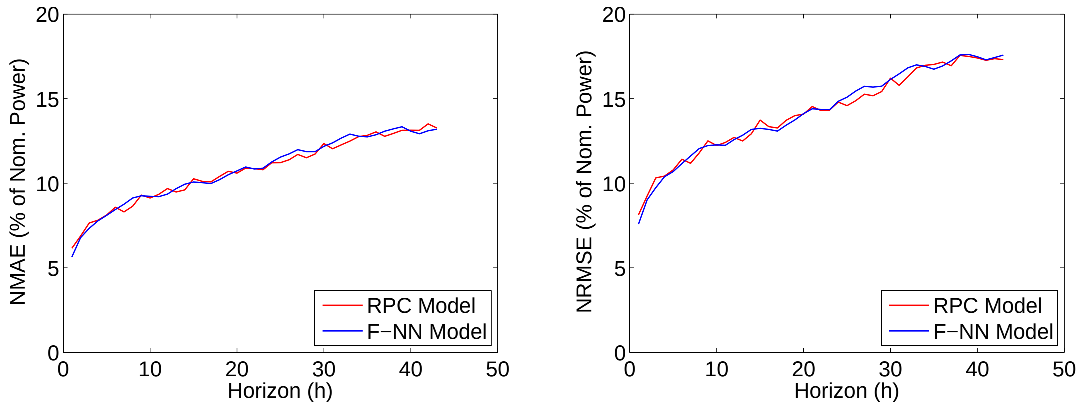
In Figure 4.17, the NMAE and the NRMSE are plotted versus the forecast horizon for the cluster approach. From a general standpoint we can note that for both criteria the performance of both model is nearly equivalent. No model clearly outperforms the other. Also, the general shape of the NMAE and NRMSE are consistent with those found in the previous case studies.

A noticeable difference in performance for both the NMAE and the NRSE criteria can be observed for the short-term horizons (1 to 6 hours ahead) when comparing the cluster approach to the direct and cascaded approaches. The cluster approach presents better performance than the direct approach and lower performance than the cascaded approach. With respect to the cascaded approach the difference can be explained by the difference in the number of on-line production measurements used as explanatory variables by both models. The cluster model examined here only benefits from the on-line measurements from one wind farm to predict a sub-area production. For the other sub-areas, the model can only rely on the NWP data. In the cascaded approach the model could rely on production measurements for all wind farms. A possible way to improve the short-term performance observed here would be to provide each sub-area model with production data from at least one wind farm from the sub-area.

With respect to the higher short-term performance of the cluster approach when compared to the direct approach, the difference can mainly be explained by the fact that the cluster approach benefits from more NWP data and from the decomposition of the total production into three separate productions. The higher amount of NWP data provides the model with a better evaluation of the short-term meteorological conditions over the entire region. Furthermore, the fact that the total production is divided into three components allows the cluster approach to more precisely model the process under consideration.

In Figure 4.18 the normalized bias of the model errors and the skewness of the error distributions are presented versus the forecast horizon. From a general standpoint the bias and skewness values do not differ greatly from those found for the direct and cascaded approaches. In addition, we can notice that the shapes of the curves as well as the values are very similar to those found for the cascaded approach.

The normalized bias curves obtained for the F-NN and RPC models are almost identical in shape to those found for the cascaded approach, the only difference being slightly



(a) NMAE

(b) NRMSE

FIGURE 4.17: *NMAE and NRMSE of the RPC and F-NN models for the cluster approach for the Irish case.*

higher values for the cluster approach. The RPC model presents positive bias for most horizons, with values very close to zero. The evolution of the RPC model bias with respect to forecast horizon is nearly constant with only a slight positive trend. The F-NN model bias presents the same characteristic shape observed for the cascaded approach: highly positive values for the first horizons with a marked, horizon dependent, decrease. Again the shape of the bias curves result from the sub-region forecast error bias curves, see Appendix B.3. In the case of the RPC model some sub-regional forecasts present negative bias while other present positive bias. When the sub-regional forecasts are summed the bias compensate each other to some extent. In the case of the F-NN model, all sub-regional forecasts present the same negative trend and most are markedly negative for higher horizons. When the F-NN sub-regional forecasts are summed there is little compensation due to the similarity of all bias values.

The skewness values found for the cluster approach are also very similar to those found for the cascaded scheme. The skewness values for both models seem to be centred around a constant value. The RPC model presents positive skewness for nearly all forecast horizons. When compared to the RPC skewness of the cascaded approach we can notice that in the present case the error distributions are less skewed for a few horizons. The skewness of the model error computed for the F-NN model also closely follows the same general shape of those found for the cascaded approach. The values for the F-NN model oscillate around zero. In the direct approach the error distributions obtained by the F-NN model were more consistently positive.

In Figure 4.19 the normalized standard deviation of the errors and the kurtosis of the horizon dependent errors are presented. Both criteria present the same shape and value ranges as those observed in the cascaded case. The NSDE values in particular are almost identical to those obtained for the cascaded case for horizons higher than six hours ahead. For short-term horizon the difference between the NSDE of the cascaded and the cluster

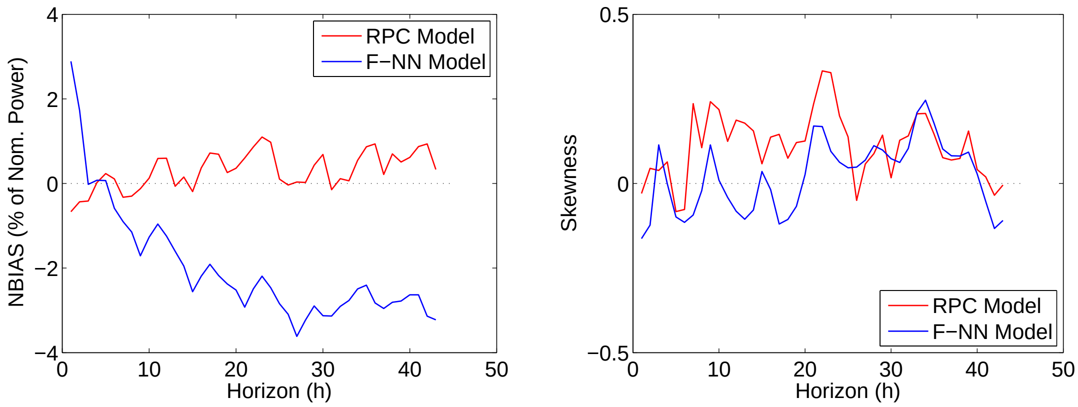


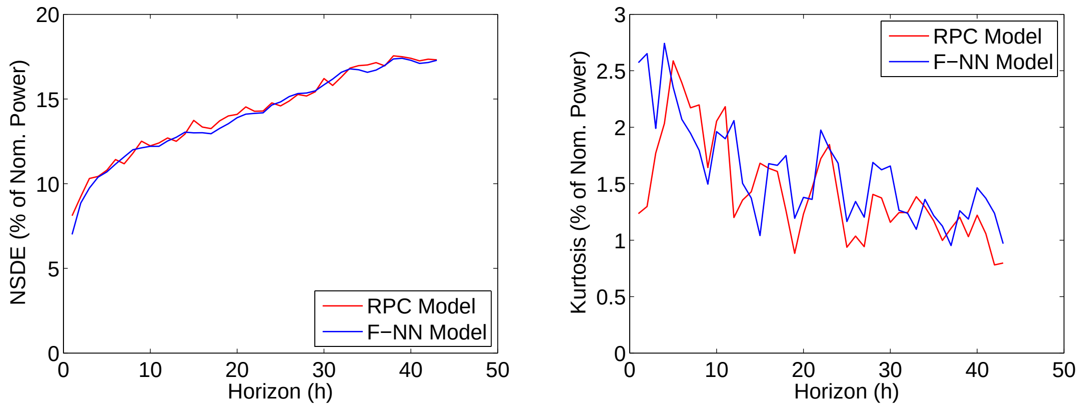
FIGURE 4.18: *Normalized Bias and Skewness of the RPC and F-NN models for the cluster approach for the Irish case.*

approaches presents the same difference as that found for the NRMSE; with a large advantage for the cascaded approach. This difference can be explained by the factors mentioned in the analysis of the NRMSE, namely fewer explanatory variables regarding current production and fewer components of the dependent variable considered as learning variables by the model.

The kurtosis values obtained in the present case are very similar for both models. When compared to the cascaded approach, the curves present the same general decreasing shape with respect to forecast horizon. We can note however that in the cascaded approach the decrease was constant over all horizons whereas in the present case the kurtosis values seem to become stable after horizon 20. This difference in behaviour can possibly be explained by the difference in the number of power measurement variables considered in both cases. In the present case, very little information concerning present production is fed into the model. As described earlier, the decreasing shape of the kurtosis is possibly due to the strong autoregressive component of regional production. Because the cluster model must rely on NWP forecasts more than on measured power the influence of power measures is only "felt" up to horizon 20. After this horizon, the kurtosis values have a behaviour that can be compared to that observed for the direct forecasting case.

In Figure 4.20 the R^2 and the improvement over OL-Persistence are presented versus forecast horizon. When compared to the curves obtained for the previous approaches we can notice that for both criteria, the results are very similar. Both the R^2 and improvement over OL-Persistence curves present the same general shape.

The R^2 values reflect the behaviour previously observed through the NMAE and NRMSE. The skill of the model is lower for the first horizons than that achieved by the cascaded approach but higher than those observed for the direct approach. This, as mentioned before, results from the number of on-line power measurements considered as explanatory variables by the cluster model. Likewise, the R^2 values found for horizons above 6 hours ahead



(a) NSDE (b) Kurtosis

FIGURE 4.19: *NSDE and Kurtosis of the RPC and F-NN models for the cluster approach for the Irish case.*

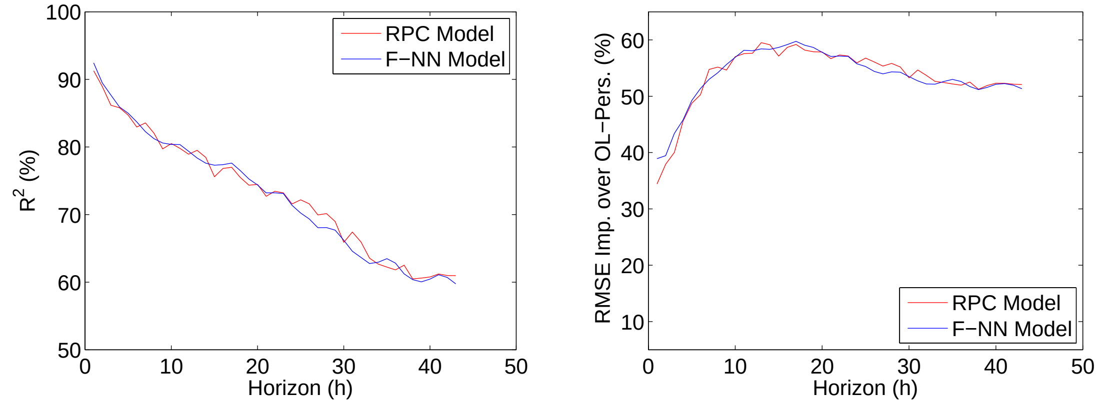
are very close to those found for the cascaded approach and superior to those found for the direct approach. When compared to each other, the RPC and F-NN models clearly present very similar performance.

The improvement over OL-Persistence is also similar for both the FNN and RPC models. For the first horizons, the improvement is rather high with values between 35% and 40% for the first horizon. This compares very favourably with the direct approach whose initial improvement ranged from 25% to 30%. Although both the direct and cluster approaches share the same on-line power measurements as past explanatory variable, the cluster approach, which exploits more future explanatory variables, can better predict the short-term production. This effect can be seen for the peak improvement, nearly 60% for the cluster approach and 57% for the direct approach.

Conclusions

The first conclusion that can be drawn from the results presented in this subsection is that the cluster approach leads to good forecasting performance when compared to the other regional approaches described in this work. One advantage over the direct approach is the computation of sub-regional forecasts that can be useful to end-users such as TSOs in the frame of network congestion management. This approach has an advantage over the cascaded approach in that it can lead to more parsimonious models in terms of the number of estimated parameters. This, in turn, can lead to better generalisation in out of sample data.

A second conclusion that can be formulated is that both the FNN and RPC models have comparable performance when used as base models in this approach. The scores obtained by both on almost all evaluation criteria are nearly identical for all forecast horizons. The only marked difference appears when examining the bias of the model errors. The bias values are comparable in absolute terms but the evolution of the bias with forecast horizon differs significantly between the two models. This difference can be linked to the difference



(a) R^2

(b) Improvement over Persistence

FIGURE 4.20: *R^2 and NRMSE improvement over OL-Persistence of the RPC and F-NN models for the cluster approach for the Irish case.*

in power curve modelling which distinguishes the models. As mentioned before, the F-NN model derives a horizon independent "power curve" whose fit quality is optimized over the ensemble of horizons, whereas the RPC model defines a different power curve for each horizon.

Further conclusions can be formulated when comparing the cluster approach to the direct and cascaded approach, by only regarding the input variables and the "granularity" of the dependent variable.

From this perspective, the cluster approach presented here outperforms the direct approach for two main reasons. First, it is clear that having NWP forecasts for more points in the region benefits the performance of the cluster approach over all horizons. Secondly, the fact that the cluster approach learns to predict the output of subgroups of wind farms (clusters) permits a finer modelling of the relation between the explanatory variables and the regional production. The improvement is higher for the short-term horizons. For longer-term horizons, the difference is not as important mostly due to the higher dependence of the power forecast on NWP forecasts and to a higher correlation of the NWP forecast errors which reduces the error smoothing effect [24].

The impact of the number of on-line power measurements used as explanatory variables as well as the "granularity" with which the regional production is learnt can also clearly be seen when comparing the cluster approach to the cascaded approach. The cascaded approach benefits from production measurements for all wind farms and models the regional production more finely by specifically training one model to forecast the production of each wind farm. The impact of these differences is evident for the short-term horizons (1 to 6 hours ahead) where the cascaded approach clearly outperforms the cluster approach. However, after the sixth horizon, the performance of both approaches is nearly identical. This is due to the use of the same NWP explanatory variables in both cases. From this, it is clear that for the short-term horizons having the most complete measure of the current production and finely modelling the relation between the explanatory variable and the dependent variable can lead to very significant performance improvements. This, however, does not seem to have a major impact on the medium and long-term performance of the approaches. Hence, for longer-term horizons, a less complex approach can be preferred over a more complex one.

4.6.5 Comparison of modelling approaches

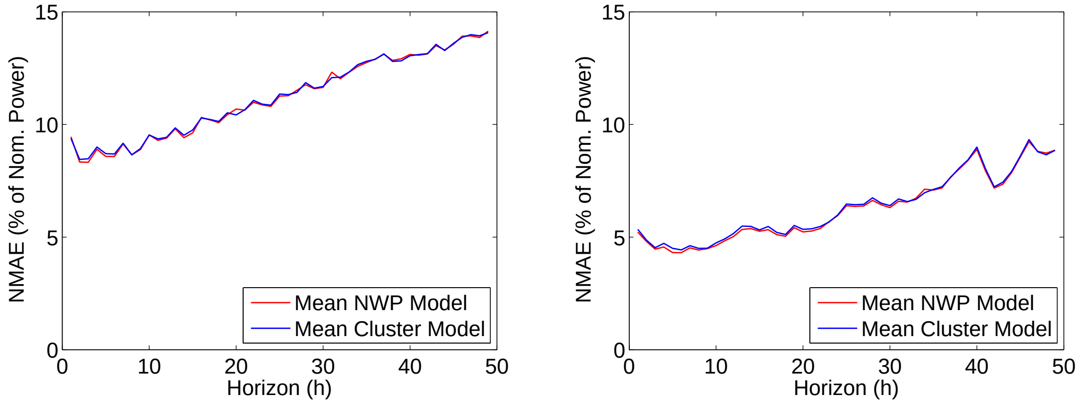
In the previous sections the aim was to compare the performance of the two base forecasting models (F-NN and RPC) in three forecasting approaches: direct, cascaded and cluster. The aim of this comparison was to establish if important performance differences exist between base models. In this section the aim is to compare the performance of two different approaches which use the same base model (namely the RPC model) and exactly the same explanatory variables, in order to establish if different approaches lead to widely different forecast performance.

Two cluster approaches, described in subsection 4.5.3, are compared in this section. The only explanatory variables used are all available wind speed forecasts. The first approach, hereafter referred to as the "Mean NWP" approach, computes an average wind speed for each cluster based on the NWP forecasts provided for the farms in that cluster. The NWP averages are then fed into a single RPC model to forecast the cluster's future production. In the second approach, hereafter referred to as "Mean Cluster", the wind speed forecasts provided for each wind farm are fed into separate RPC models, which compute forecasts of the production of the cluster where the reference wind farms are located. The different forecasts computed for a given cluster are then averaged to obtain a single forecast for the cluster. In both approaches the cluster forecasts are simply summed to obtain the regional forecast.

The main difference between the two approaches is the location of the averaging procedure in the model chain. This leads to a significant difference in the dimension of the models. In the Mean NWP approach only three RPC models are tuned whereas in the Mean Cluster approach N RPC models are tuned, N being the number of wind farms for which NWPs are provided or more generally the number of NWP grid points for which wind speed forecasts are available.

The results presented here are for the Irish and Danish case studies. In the Danish case we only consider the forecast of the aggregate production of the 23 reference wind farms for which detailed data was available. This evaluation is different from the direct approach presented in Appendix B.2, where the total production of the Jutland-Funen area was considered. For the sake of brevity only the NMAE, R^2 and NBIAS criteria are discussed in detail in the following paragraphs; the complete set of evaluation criteria can be found in Appendix B.4.

In Figure 4.21 the NMAE versus forecast horizon is presented for both cases. The first aspect that we can notice is the near equality of the NMAE values obtained for both forecasting approaches. In both cases the difference between NMAE values is very slight for



(a) Ireland (b) Denmark

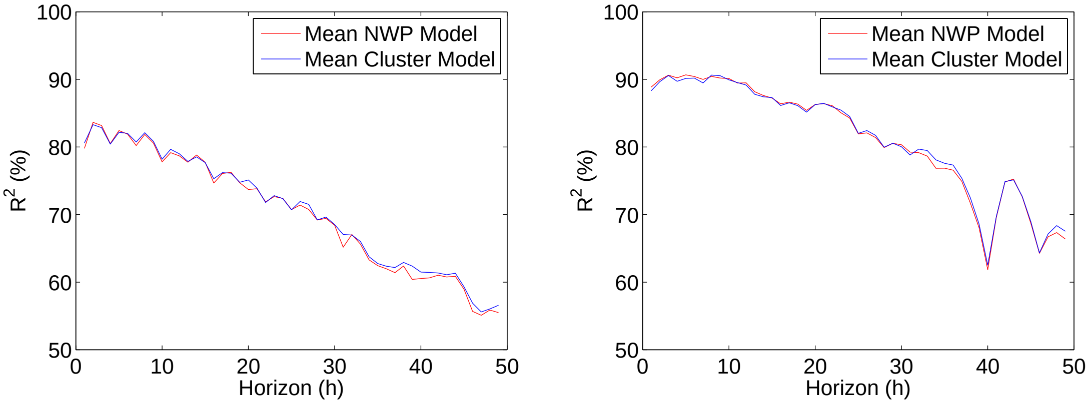
FIGURE 4.21: *Normalized mean absolute error versus forecast horizon for the case of Ireland and Denmark using the mean meteorological approach and the mean cluster approach*

most horizons. The difference between both approaches is most apparent in the Danish case with a very slight advantage for the Mean NWP approach.

A second aspect that can be noted is the shape of the NMAE curves for the first forecast horizons. In the cases presented so far, the NMAE curves presented low values for the first horizon with a sharp increase over the next horizons. Here, the initial behaviour of the NMAE is radically different. The NMAE values over the first horizons decrease and attain minimum values for horizons between 3 and 6 hours ahead depending on the case study. This difference can be explained by the absence of on-line measurements as explanatory variables. This behaviour is typical of models that only rely on NWP data, and can be explained by the fact that NWP forecasts often present higher accuracy for medium-term horizons than for very short-term horizons.

When comparing the results of the Irish case to those obtained for the cascaded and cluster approaches presented earlier, we can notice that, aside from the initial difference, the performance of the Mean NWP and Mean Cluster approaches is comparable for horizons above 10 hours ahead. This is understandable given the importance of forecast wind speeds as explanatory variables when computing power forecast for medium and long-term horizons. When compared specifically to the results obtained when using the RPC model as a base model in the previous approaches, the same NMAE pattern is visible in all cases. This illustrates the strong horizon dependent homogeneity of the NWP forecasts provided for the different farms. This behaviour does not change even when the NWPs are averaged.

The horizon-dependent evolution of the NMAE in the Danish case presents two noticeable aspects. There seems to be a 12 hour cycle in the NMAE values, with slight peaks at 12, 24 and 36 hours ahead. The second aspect appears after horizon 36. It is a sharp cyclical variation of the NMAE values, with a 6-hour period. When looking back at the NWP data provided by the Hirlam NWP model in this case, it appears that the wind speed values provided between horizons 36, 42 and 48 hours are the result of a linear interpolation be-



(a) Ireland (b) Denmark

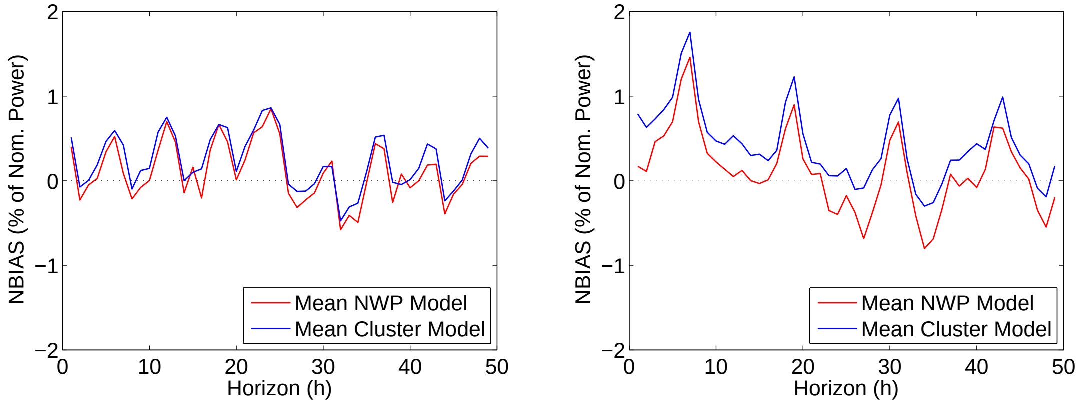
FIGURE 4.22: *R^2 versus forecast horizon for the case of Ireland and Denmark using the mean meteorological approach and the mean cluster approach*

tween the values computed for horizons 36, 42 and 48. This can possibly explain the NMAE variations observed for horizons above 36 hours.

In Figure 4.22, the R^2 criterion is presented versus forecast horizon for both cases. We remind the reader that although the R^2 criterion is computed here as an improvement over the climatological mean production forecast, it captures the variance of the time series explained by the forecasts. Here, we can notice that the values obtained are again very similar for both approaches. Also, the shape of the R^2 curves differs for the first horizons from those found in previous cases. As with the NMAE values, for medium and long-term horizons, the R^2 values found for the Irish case are very similar to those found for previous cases. Also, in the Danish case, the odd variations observed for the NMAE for horizons above 36 hours ahead, are present.

The absence of production measurements as explanatory variables leads to horizon dependent R^2 curves whose shape only depends on the quality of the NWP forecasts and the ability of the model to capture the relation between wind speed forecasts and wind power production. We can notice that the shape and the variations of the R^2 curves obtained here closely match those of mutual information curves presented previously in Figure 3.16

The close correspondence between the mutual information curves and the R^2 curves reflects the ability of the approaches examined here to model the relation between the forecast wind speed variables and the regional power. Indeed, a perfect model will perfectly translate the relation between explanatory and dependent variables. A less-than-perfect model will not accurately capture the relation and will provide a distorted approximation. The R^2 criterion and the mutual information both measure the strength of the dependence between two variables, albeit from different perspectives. In the present case, the R^2 criterion measures the relation between the forecasts, in other words the dependent variables "translated" into power, and the measured power. The mutual information directly measures the strength of the relation between the explanatory and dependent variables. The



(a) Ireland

(b) Denmark

FIGURE 4.23: *Normalized bias versus forecast horizon for the case of Ireland and Denmark using the mean meteorological approach and the mean cluster approach*

similarity in the strength of dependence, as measured by both criteria, can be seen as resulting from a good "translation" of the dependent variables into power forecasts by the two approaches. In other words, the translation performed by the models did not to strongly distort the true relation that exists between the explanatory and dependent variables.

Furthermore, it can be noticed that the mutual information values are lower for the Irish case than for the Danish case. Although this does not necessarily imply that the relation is weaker for the Irish case, the same difference can be observed for the R^2 values. Also, when comparing the mutual information between individual wind speed forecasts and regional production (see Figure 3.15) we can notice that the values for the Irish case are far lower than those obtained for the Danish case. This would seem to indicate that, aside from different smoothing benefits, the lower performance observed in the Irish case might be linked to less "regionally representative" NWP forecasts. This possibly results from the higher terrain complexity of the Irish case.

From this remark we can further suppose that although the models might not be perfect, an important part of the forecast error results from the lack of information or of explanatory strength of the input variables used by the models. From this, it appears that concentrating on improving the model's fitness capacity is important, but, given the quality of existing models, it might be more beneficial to investigate the improvement provided by carefully selecting explanatory variables.

In Figure 4.23 the normalized forecast bias versus forecast horizon is presented for both cases. This criterion can be seen as giving a measure of the information remaining in the residuals not captured by the model. In this sense, the two approaches do not perfectly model the relation between the explanatory and dependent variables. However, the magnitude of the bias is relatively small for both cases, and is comparable to the biases found for the other regional forecasting approaches. The evolution of the NBIAS values for both cases is also similar to those observed in the previously examined cases.

Closer inspection of the NBIAS values does seem to indicate that the Mean NWP model is slightly less biased than the Mean Cluster model. To investigate this we compute the mean absolute bias over all horizons. For the Irish case the mean absolute bias equals 0.28% of nominal power for the Mean NWP model and 0.31% for the Mean Cluster model. In the Danish case the mean absolute bias equals 0.35% for the Mean NWP model and 0.47% for the Mean Cluster model. These results do indicate that in the two cases presented here the Mean NWP model is less biased than the Mean Cluster model. However, it must be noted that the difference is very slight, in the order of a tenth of a percentage point, so definitely concluding on the clear superiority of one model over the other is not possible.

Conclusions

The first aspect that must be noted is that there is no clear performance difference between the two approaches examined here. In the present case different regional forecasting approaches that use the same explanatory and learning variables lead to almost identical forecasting performance. Only a slight "superiority" of the Mean NWP approach is apparent in the evaluation.

The results clearly show the important benefit that can be derived from the use of on-line power measurements as explanatory variables. In the present case no power measurements were used and the forecast performance for the first horizons is clearly inferior to that observed for previous approaches where on-line power data was used.

The approaches' performance in terms of modelling capacity seem similar and adequate as shown by the similarity between the R^2 and mutual information curves. From this, it can be concluded that the forecast error attributable to modelling errors is relatively small. It can be supposed that a non-negligible part of the forecast error is due to lack of accuracy of the wind speed forecasts, which are the basis of the power forecasts. This remark can further be extended to the other explanatory variables when considering the results presented for the previous approaches. It appears that the observed performance for all cases can, to a large extent, be explained by examining the explanatory variables used to compute the forecasts. Therefore, adequate selection of explanatory variables is essential in order to maximize forecast accuracy.

4.7 General conclusions

In this chapter we have presented the regional forecasting problem from a statistical point of view. In this sense we have described the problem as a learning problem where a model must be determined from historical data. This data comprises the regional wind power and time series of explanatory variables, which can be the regional power time series itself, aggregates of regional wind power in the form of time series of individual wind farm production and "time series" of NWP forecasts.

From this perspective, we have seen that existing base models can be combined in different ways to compute regional production forecasts. We have identified three different generic combination approaches. These different approaches can be used alone or can be further combined in an infinite number of ways, depending on the characteristics of the base models. It is also possible to combine different base models in a single approach. Given the large number of different configurations that can be derived from these approaches, two questions arise. The first is: to what extent is forecast accuracy linked to the base model used in a given model combination? The second question is: to what extent is forecast accuracy linked to the way in which base models are combined?

To answer these questions we have compared the performance of different combinations based on two models (F-NN, RPC) whose performance has been shown to be in line with that of other state-of-the-art models. By examining the results of these combinations on two different case studies we have gained some insight on the answers to the above questions.

The answer to the first questions is that differences do exist, but the difference between base models is very slight in the case of the presented approaches. Also, from one region to the other, no clear superiority was found. This agrees with the results of other competitions such as the Anemos competition [32] and the M-3 competition [115]. Indeed, when comparing advanced forecasting models that have high modelling capacity, and whose learning mechanism ensures good approximation, only minimal performance differences are observed. This conclusion is obviously partial. To effectively determine a statistically significant difference, results on a statistically significant number of case studies would have to be available. Such results would allow statistical testing methods to be applied as was the case for the M-3 competition [116].

The answer to the second question is quite similar to the answer to the first question. Indeed, the same behaviour was witnessed when comparing different forecasting approaches, which use the same input data and the same base models to provide the forecasts. No clear difference is observed. From this we can deduce that it is probable that as long as the base model used in an approach provides reasonable modelling behaviour and the forecasting approach is reasonable, then good performance is likely. In other words, the type of approach should, in our sense, answer a more pragmatic requirement that is: provide the forecasts useful to the intended end-user. Indeed, based on the needs of the end-user and the availability of explanatory data (NWP forecasts, measured data, existing single wind farm forecasts), a reasonable combination can almost always be found. One aspect that should be kept in mind is to limit, as much as possible, the complexity and the dimensionality of the model combination when seen as a whole regional forecasting model.

In the same perspective, it appears that forecast performance depends mostly on the variables used as explanatory inputs to the models. This is of course the case in all forecasting fields. In particular, the wind power forecasting community recognizes that an important part of the power forecast error results from weather forecast errors. Hence, care must be taken in order to choose the variables that lead to the best forecast accuracy. In the case of regional forecasting, the problem is exacerbated to a certain degree given that many more variables are available to the forecaster. From the comparison of the results obtained here for the different approaches, it appears clearly that having no on-line power data highly penalizes forecast accuracy for the first horizons. Also, having all power measurements proves useful, but not necessarily much more so than having only power measurements for a few farms. The same can be observed for NWP data. Using wind speed forecasts for more than one location is better than only using those from a single location. However, the use of all available locations does not significantly improve the medium and long-term accuracy of the power forecast. It appears that to perceptibly improve forecast performance for regional forecasting, adequately selecting the most relevant explanatory variables can be more profitable than trying to obtain the model that has the best generalisation properties.

From the results obtained here it can also be concluded that, obviously enough, wind speed forecasts are the most important explanatory variable when predicting wind power. On-line power measurements also provide very useful information for the short-term forecast horizons. Further, forecasting models should be "horizon conscious" in order to explicitly model the horizon-dependent relation between the explanatory and dependent variables. Other meteorological data, such as wind direction forecasts, can prove useful in specific cases, for example Alaiz wind farm which is located in very complex terrain, has a very direction dependent "power curve". Modellers should, however, keep in mind that there often exists a strong statistical dependence between different meteorological variables. Hence, adding them as explanatory variables will not necessarily provide the model with more information on the future power production. Adding mutually dependent variables to the list of explanatory variables will however increase the dimensionality of the problem. This can hinder statistical models by exposing them to the curse of dimensionality, which lurks behind all statistical learning theory.

CHAPTER 5

Explanatory Variable Selection

Abstract

This chapter examines the impact of explanatory variable selection on forecast accuracy. The first part proposes a methodology for studying the influence of different explanatory variable subsets on regional wind power forecast accuracy. This methodology is then applied to measured production time series and to wind speed forecasts. From the detailed analysis of the results, general conclusions are derived and the necessity of performing variable selection is clearly identified. This leads to the second part of the chapter, which investigates the application of variable selection techniques to the regional wind power forecasting problem. This part starts with an overview of existing variable selection methodologies. From this review three methods, based on different methodologies, are proposed. One is a state-of-the-art filter method taken as an "off the shelf" method. The two other methods, a clustering-based filter algorithm and a stability-based wrapper algorithm, are proposed here. These algorithms exploit certain characteristics of the regional wind power forecasting problem and result from the adaptation of existing methodologies. The three methods are benchmarked on the Irish and Danish case studies. From this benchmark a detailed analysis of the properties of the methods is carried out and general conclusions on their usefulness are provided.

5.1 Introduction

In the previous chapter we examined the performance of different regional forecasting models for different test cases. Some differences in performance were shown to be attributable to the models. However, the most important differences appeared to be due to differences in the explanatory variables used as input to the models. The influence of explanatory variables is not surprising and has long since been recognized as a major performance factor in all forecasting fields. Indeed, in physics, when establishing a model one must identify the physical factors that govern the process under consideration; the weight of the apple is more strongly conditioned by its mass than by its colour. The same holds true in statistical forecasting. Indeed, in classical time series approaches much effort is devoted to selecting the right explanatory variables that must be derived from the time series under consideration.

Hence, the question which must be answered is: which explanatory variables should be used in order to achieve the best[1] possible forecasts? This question is vast and can be considered from different perspectives. From a wind power forecasting standpoint, the question can be reformulated as: which physical variables are involved in the power production process? As seen in previous chapters, wind speed as well as wind direction play an important role in the power conversion process. However, these variables are not the only ones influencing the process. From a more statistical perspective, the question can be rephrased as: which variables are the most explanatory given the inherent uncertainty they contain and the dependencies existing between them? This aspect relates to the information content of the variables and their possible redundancy, which can hamper statistical learning models. In the frame of regional forecasting, this question becomes more important given the higher number of highly dependent explanatory variables available to the forecaster. In this case, variable selection can be seen as being the search for the variable subset that leads to the best forecasts.

In this chapter we will examine the problem of choosing the explanatory variables that will lead to the best forecasts. To do so, we will first discuss the different aspects of variable selection in a regional wind power forecasting context. We will then focus on one of these aspects: the selection of the most useful explanatory variable subset from a given set of explanatory variables. In order to examine this problem we will proceed in two steps. First we will examine the impact that different explanatory variable subsets can have on forecast performance. Having shown the benefits that can be derived from adequately selecting the variable subset used by a regional wind power forecasting model, we will then examine the possibility of implementing explanatory variable selection methods in the frame of regional wind power forecasting. Indeed, in this context variable selection is complicated by the fact that the considered variables can all be considered relevant and somewhat redundant.

The aim of the subset performance framework that we propose is twofold. The first aim is to determine to what extent different variable subsets can lead to different forecast accuracy. The second aim of the framework is to determine variable subset characteristics that can be used to facilitate the selection of a variable subset that leads to the best, or at least to good, forecast performance. The characteristics that will be examined include, among others, subset cardinality, reference wind farm properties, and geographical distribution of the explanatory variables.

Having shown the benefits that can be derived from variable subset selection, we will investigate the possibility of implementing variable selection methods in the frame of re-

[1] Here, the best forecasts are those that present the best characteristics in terms of the evaluation criteria presented in section 2.3.

gional wind power forecasting. From a state of the art review and using the findings of the subset performance framework, we will propose two selection methods that will be benchmarked against a state-of-the-art method. The performance of these three methods will be analyzed on two cases studies. From this analysis the relative merit of each method will be discussed, and we will show to what extent these methods can help modellers in their variable selection task.

5.2 Explanatory Variables Selection for Regional Forecasting

The question of what explanatory variables to use in regional forecasting has not been explicitly investigated in the regional forecasting literature. From the description of some regional forecasting applications, it appears that forecasters have concentrated on using NWP variables and production measurements traditionally used for single wind farm forecasting [24, 90, 93]. With respect to the number of variables to be used or to the number of reference farms that are necessary to attain good forecasts, the approach has been mainly to use a high number of variables. For example, in [24] the authors suggest that the data from 40 to 50 reference wind farms is sufficient to adequately forecast the regional production of a region the size of Germany. In [93], the authors use wind speed forecasts provided on a 55 km x 33km grid covering all of Germany, to compute the total German wind power production. However, in this example the authors use principal component analysis in order to reduce the dimensionality of the model.

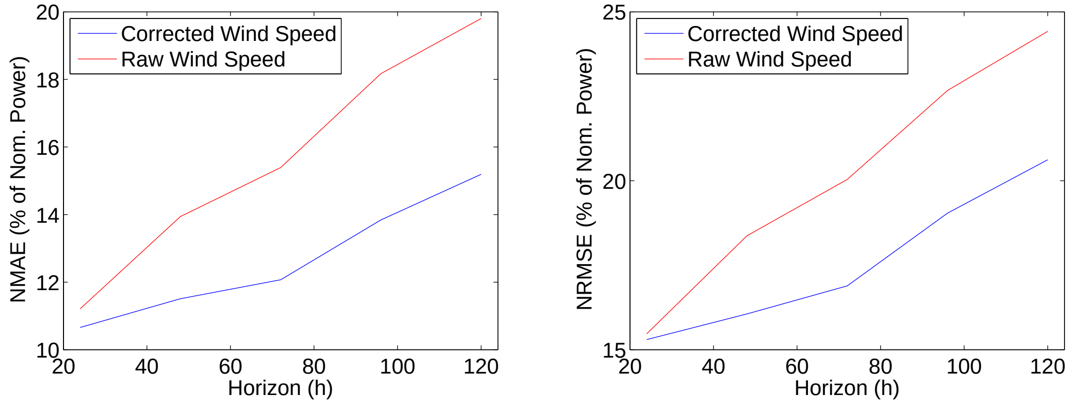
From these elements, the question of what explanatory variables should be used for regional wind power forecasting can be divided into two different questions : which variables should be used? and how many variables should be considered?

Which variables for regional forecasting?

This question can be considered from a theoretical and a more practical perspective. We will shortly discuss both of them.

A first answer to this question can largely be provided by the research that has been conducted on modelling the atmosphere and the wind to power conversion process [28, 29]. The physical models of wind energy conversion systems are well established and clearly point to the variables that have the most influence on wind turbine power output. With this knowledge it is clear that forecasters will not attempt to use statistical models to forecast wind power based on temperature and total cloud amount forecasts alone. Although these variables can be linked to wind power production, the choice wind speed forecasts as explanatory variables will evidently lead to much more accurate results.

Furthermore, it can be shown that considering different variables can help improve forecasts for all horizons. To illustrate this point we can consider the results presented in Figure 5.1, which are from [67]. The NMAE and NRMSE in the figure were obtained using the F-NN model with two different wind speed forecasts: “raw” wind speed forecasts and forecasts corrected for bias using Kalman filters and wind speed measurements. From this fig-



(a) NMAE (b) NBIAS

FIGURE 5.1: *Normalized mean absolute error and normalized root mean square error for Rokas wind farm, located on the Greek island of Crete, using raw and filtered wind speed forecasts [67].*

ure it is clear that the filtered wind speed forecasts lead to better wind power forecasts. This can directly be linked to the improvement in accuracy of the filtered wind speed forecasts.

From a practical, or operational, perspective, the answer can be considered from the point of view of data quality. In the literature, definitions of data quality are varied and all do not necessarily comprise the same aspects [117]. Data quality can be defined along several dimensions, the most commonly sited ones being: *accuracy*, *timeliness and currency*, *completeness* and *consistency* [118]. Definitions of each of these dimensions vary. Accuracy should generally be understood in its common conception: the recorded value reflects the real state of the considered variable. Accuracy can in our case refer to the accuracy of wind speed measurements (i.e. how accurate is the anemometer) or to the accuracy of a NWP model's forecast of the future wind speed value. Timeliness refers to the availability of the data, in the present case this would translate to measured and forecast variables being provided on time so that power forecasts can be delivered when needed by the end-user. Currency of the data translates to the time when a data item was stored, in other words this can be linked to the frequency of measurements and to the time steps for which forecast variables are provided. Completeness refers to the availability of all values. In our case this translates to availability of all the values of the considered variables. This can be particularly important during statistical learning where increasing training set size often translates into more robust model tuning. Finally, data consistency refers to data values being the same for the same situation. In the wind forecasting case this means that changes to measuring tools as well as changes in forecasting models should be clearly noted. Changes in measurement precision or NWP model accuracy in a training set can lead to degraded performance. Such changes should be handled specifically.

These quality criteria for the variables are important especially if the forecaster is given a choice when selecting different representations of the same variable. This is particularly true in an operational setting where wind power forecasts are used as explanatory or decision variables in decision-making processes related to trading or network regulation. In this setting, wind power forecasts are subject to the same quality requirements as the explanatory variables used in the forecast generating process. However, selecting data based on its quality is seldom possible in practice. When dealing with measured variables, different measurements of the same variable are not common in wind power applications. The case when selection might be possible is when selecting between different NWP providers that run different models. In this case all the dimensions are important, but the one that has the most impact on wind power forecast performance is accuracy.

Although both the "physical" and "practical" aspects of this question are important and should be considered when implementing an operational forecasting system, providing a definitive answer to the question which explanatory variables should be used in regional forecasting is outside the scope of this thesis. Here we concentrate on answering the second question: how many explanatory variables should be considered to provide accurate regional wind power forecasts.

How many variables should be considered?

In the frame of single wind farm forecasting, power measurements and NWP forecasts of different variables (wind speed, wind direction, temperature, etc.) and for different altitudes are usually considered. This usually translates to at least three variables (power measurement and forecasts of wind speed and direction) for each wind farm, although in general more variables are available. When transposed to the regional scale, where several reference wind farms are available, this can translate to several tens or hundreds of variables. The problem is then to select the variable subset that leads to the best forecast performance.

Finding a solution to this problem is important for at least two reasons. First, from an operational perspective, limiting the number of explanatory variables reduces the complexity and cost of the forecasting system. Secondly, training statistical models that consider many explanatory variables can be problematic and lead to "sub-optimal" performance. Although the operational perspective is of interest, it is outside the scope of this work. Here we will concentrate on the performance issues that may arise when statistical models consider too many variables.

The Kalman input filtering example presented above clearly illustrates the important improvement that can result from using several explanatory variables. However, increasing the number of explanatory variables can also hamper statistical models. Having too many explanatory variables can lead to what is generally referred to as the *curse of dimensionality*. This term was first used by Bellman [119] in the field of combinatorial optimization, where the computational cost of solving problems was observed to grow exponentially with the number of problem dimensions. In statistics the term has been used to describe the increasing sparsity of data in spaces of increasing dimension. This effect can easily be illustrated by considering a hypercube and the inscribed hypersphere. If we consider the hypercube $[-1, 1]^d$ where d denotes the dimension of the hypercube, and the inscribed hypersphere of radius $r = 1$, the ratio between the volume of the hypersphere and the volume of the hypercube decreases exponentially with the dimension [100]. The values of this ratio

TABLE 5.1: *Fraction of the volume of a hypercube* $[-1,1]^d$ *lying in the inscribed hypersphere and the quadrant hypercube* $[0,1]^d$ *[100].*

Dimension d	1	2	3	4	5	6	7
Frac. in Hypersphere	1	0.785	0.524	0.308	0.164	0.081	0.037
Frac. in Quadrant Hypercube	0.5	0.25	0.125	0.062	0.031	0.016	0.007

are given in Table 5.1 for low dimensions. From this it is clear that, considering a uniform sample of points within the hypercube, almost no points will be contained within the hypersphere. Even the large inscribed hypercube $[0,1]^d$ will only contain 2^{-d} of the data points. Hence, in higher dimensions a "local" neighbourhood of the space will almost certainly be empty, whereas if the neighbourhood is not empty then it is almost certainly not "local".

From a statistical perspective, the number of explanatory variables considered by a model conditions the dimension of the input space. The dimension of the input space is inferior or equal to the number of explanatory variables, with equality when the explanatory variables are not collinear. The sparsity of data poses a problem when estimating model parameters. In order to obtain good parameter estimates, a statistically significant number of observations must be available to the estimation procedure. In high dimensional input spaces this can be achieved in two ways. Either consider observations belonging to a very large neighbourhood, in which case the local behaviour of the function being modelled is lost, or ensure that an astronomical number of observations are available in the dataset.

This clearly illustrates the fact that in order to accurately model a function, the number of available observations in the training set conditions the maximum number of explanatory variables that can be considered by a model. For a given estimation accuracy it is possible to compute equivalent sample sizes for different dimensions. In the frame of multivariate density estimation, Scott presents equivalent sample sizes for the estimation of multivariate Normal distributions [100]. For an initial 50 observations in dimension one, more than one million observations are necessary in dimension 8 in order to achieve estimates of equal accuracy [100].

The estimation of the model parameters can be further complicated when noise, or uncertainty, is present in the explanatory variables. Indeed, if the explanatory variables are uncertain, or noisy, accurately estimating the model parameters requires a greater number of observations than if the variables were perfectly certain or noiseless. Here an explanatory variable is considered as uncertain, or noisy, if the conditional density of the dependent variable given the explanatory variable is not a Dirac density for all possible values of the dependent variable.

In this chapter we will investigate the performance that can be expected when different explanatory variable subsets are considered. From this investigation we will show that "sub-optimal" performance can occur when considering different explanatory variable subsets in regional wind power forecasting. In this examination we will concentrate on determining explanatory variable subset characteristics that can prove useful when searching for subsets that are likely to lead to the best performance. After this investigation, we will explore the possibility of implementing selection algorithms that take into account the specific nature of the explanatory variables encountered in regional wind power forecasting and that exploit the variable subset characteristics revealed in the initial investigation.

5.3 A Framework for Examining the Impact of Variable Subset Selection on Regional Forecasting Accuracy

As we have seen in the previous section, both the quality and the number of explanatory variables considered by a model condition its accuracy. In this section we present the methodology that was followed in order to investigate the relation between input variable subsets and regional forecast accuracy. Our aim is first to evaluate the performance obtained for all variable subsets using a given forecasting model. Then, we will investigate the characteristics of the variable subsets that lead to the best performance.

To meet the first objective we first determine the total number of subsets that must be evaluated. We then propose a regional forecasting model that can provide reasonably accurate forecasts and can meet the computational constraints of the evaluation. Finally, we propose a criterion for evaluating the forecast performance derived from the use of different explanatory variable subsets. To meet the second objective we define the subset characteristics that will be investigated. Establishing a link between the variable subset characteristics and the performance of the subsets can prove useful in the definition of efficient methods for explanatory variable selection.

5.3.1 Determining the Number of Explanatory Variables Subsets

The aim here is to examine the characteristics of the variable subsets as they relate to the performance of regional wind power forecasting models. For this purpose, the forecasts resulting from the use of each variable subset must be evaluated. This means that a regional forecasting model has to be run using as input each possible explanatory variable subset. As will be shown here, performing such an evaluation is far from trivial.

Assuming that the order in which the variables are submitted to the forecasting model is irrelevant to forecast accuracy then, there are C_k^n possible subsets of k variables chosen from a set of n variables. The number of combinations C_k^n of k elements chosen among n is defined as [120]:

$$C_k^n = \frac{n!}{k!(n-k)!}, \tag{5.1}$$

where ! denotes the factorial function. From this, the total number of variable subsets which can be derived from a set of n variables is $\sum_{k=1}^{n} C_k^n$. The binomial theorem states that [120]:

$$(x+y)^n = \sum_{k=0}^{n} C_k^n \, x^k y^{n-k}, \tag{5.2}$$

where x and y are real, and n a non-negative integer. The number of subsets can be derived by setting $x = y = 1$, then

$$(1+1)^n = C_0^n 1^0 1^n + \sum_{k=1}^{n} C_k^n 1^k 1^{n-k}. \tag{5.3}$$

Given that $C_0^n = 1$, $1^0 1^n = 1$ and $1^k 1^{n-k} = 1$ for all integer values of n and k, we have:

$$(1+1)^n = 1 + \sum_{k=1}^{n} C_k^n, \tag{5.4}$$

and therefore the total number of subsets is:

$$\sum_{k=1}^{n} C_k^n = 2^n - 1. \tag{5.5}$$

From this, it is clear that considering a large set of explanatory variables will lead to a very large number of subsets to be examined. As mentioned earlier the number of explanatory variables available in a regional forecasting context is important. In the case studies presented in section 3.5, 143 "raw" explanatory variables are available for the Irish case and 414 for the Danish case. By "raw" we mean explanatory variables as provided by the end-users and by NWP forecast providers; supplementary variables can be created by combining these variables. Considering only raw variables would lead to the evaluation of $2^{143} - 1 \approx 1.11 \times 10^{43}$ variable subsets in the Irish case and $2^{414} - 1 \approx 4.23 \times 10^{124}$ in the Danish case. Clearly, investigating the forecasts derived from such a large number of variable subsets is intractable, if not impossible.

Investigating variable subsets confronts us to the *curse of dimensionality*, but from a combinatorial optimization perspective. The question of which variable subsets lead to the best forecast accuracy remains, but it is virtually impossible to answer if too many variables are considered. We therefore need to: use a forecasting model that possesses low computation time, and limit the number of explanatory variables considered.

5.3.2 The Considered Regional Forecasting Model

We have determined the number of variable subsets that must be evaluated. We also showed that in the frame of this study, the considered regional wind power forecasting model must possess short computation time. Furthermore, it must also have the ability to handle multi-dimensional inputs. Here, we analyze the implications of these two requirements and we propose a model that meats them.

In the literature, many models capable of exploiting multiple explanatory variables exist. These range from ARMAX models to neural networks and support vector machines. Selecting a model from the vast corpus of available models is not easy; especially since most of them can be shown to be conceptually very similar [114]. Also, different modelling and learning approaches have been shown to present different behaviours when faced with irrelevant or redundant explanatory variables. In a nutshell, a variable is relevant if it provides the model with useful information on the modelled process. Clearly, irrelevant variables can only disrupt a model's learning capacity. Redundant variables are variables that provide the same, hence redundant, information on the process being modelled. Relevancy and redundancy of explanatory variables will be discussed in greater detail in subsection 5.5.1. The accuracy of some induction tree algorithms has been shown to degrade when confronted with irrelevant explanatory variables [121]. The accuracy of other models, such as Bayesian classifiers, has been shown to be penalized when dealing with redundant explanatory variables [122, 123]. From the literature, no clear overall advantage can be attributed to any type of model. Therefore, in order to select a regional forecasting model for the purpose of this work we will look at models that have been proposed in the literature to provide regional wind power forecasts.

In [85] neural networks are used to compute single wind farm forecasts that are then aggregated. In [84] conditional parametric models are used to provide multiple regional forecasts, which are then linearly combined to provide the final forecast. In [124] the team responsible for the development of the Previento forecasting system [57] reports using an upscaling method based on linear regression. Finally, the model described in [58] is based on the aggregation of single wind farm forecasts computed using a CFD-based model. From the above references, it appears that models based on single wind farm physical forecasting models use some sort of aggregation model. Statistical models merge several single wind farm, or local or total forecasts to provide the final regional forecast. The combination is linear in the model described in [84]. Finally, single wind farm forecast combination models, such as the one described in [125], are also linear in nature. From this, it appears reasonable to use a cascaded approach (see section 4.3) with a linear aggregation of single wind farm forecasts to perform the upscaling step.

Using a linear cascaded approach is not only similar to existing state-of-the-art regional forecasting approaches but it also offers the possibility of combining single wind farm forecasts provided by different forecasting models. In this way, any type of single wind farm forecasting model can be used as the base model in a regional forecasting application. Another advantage of such an aggregation method is that it is relatively inexpensive from a computational point of view. Indeed, as remarked in the previous section, the aim here is to examine several thousands of combinations. Using linear aggregation permits the use of advanced methods to model the nonlinearities existing between explanatory variables such as wind speed and wind farm production. The single wind farm forecasts can be computed once, and then only the aggregation model needs to be run thousands of times in order to compute the results for each explanatory variable combination. In this way the CPU-intensive nonlinear modelling step is performed only once, greatly reducing the computational burden of the evaluation.

Finally, from a more physical point of view, a linear model can be justified by considering the nature of regional production. As mentioned in section 3.2, regional production is the sum of the productions of the wind farms located in a defined region. Hence, if a single wind farm forecasting model is tuned for each wind farm within the region, summing all the forecasts should lead to a reasonable estimate of the future regional production. Furthermore, if the single wind farm forecasts errors are random variables, and independent from each other, then summing the forecasts is expected to lead to a reduction of the regional forecast error through statistical smoothing.

Although using a linear model ensures fast computation of the regional forecasts, parameter estimation must also be rapid in light of the numerous executions that will be required. The aggregation model must be capable of performing the upscaling or extrapolation from the single wind farm productions to the total production. In order to fulfil these requirements we propose the use of least squares linear regression as the parameter estimation method. The regional prediction model for horizon k can thus be defined as:

$$f_k^{reg}(F_{S,k}^{wf}) = \alpha_{0,k} + \sum_{i \in S} \alpha_{i,k} f_{i,k}^{wf}(x_{i,k}) \tag{5.6}$$

where $S \subset [1, L]$ is the subset of wind farm indexes considered by the model, L the number of reference wind farms available, $F_{S,k}^{wf}$ the function vector of single wind farm models $f_{i,k}^{wf}$ for horizon k, tuned for the wind farms whose indexes are in S, $x_{i,k}$ the explanatory variables associated to the reference wind farms whose indexes are in S and $\alpha_{0,k}$ and $\alpha_{i,k}$ are the regression parameters. The regional power forecast is then:

$$\hat{P}^{reg}(t + k \mid t) = \alpha_{0,k} + \sum_{i \in S} \alpha_{i,k} w \hat{P}_i^{wf}(t + k \mid t) \tag{5.7}$$

where $\hat{P}^{reg}(t + k \mid t)$ is the regional power forecast for horizon $t + k$ computed at time t and $\hat{P}_i^{wf}(t + k \mid t)$ are the power forecasts computed at time t for horizon $t + k$ for the wind farms whose indexes are in S. The regression parameters in Equation 5.7 are computed using the least squares multiple regression method described in subsection 4.4.2.

5.3.3 Determining the forecast performance of an explanatory variable subset

When examining which variables lead to the best out-of-sample forecasts, the objective is to find a single explanatory variable subset that provides the best results. In order to carry out this “selection” however, it is necessary to define the criteria used to evaluate the performance obtained using a given subset. The criteria that will be used here are the ones defined in section 2.3. More specifically, we will concentrate on the NRMSE and NMAE criteria. The NRMSE is linked to the sum of square errors; in fact it corresponds to the criterion that is optimized by the linear least squares regression performed by the regional forecasting model. This will be the criterion that will be used to determine the best subsets.

In the literature, explanatory variable subsets are examined considering only single horizon forecasts. In the present case the forecasts are provided for several time steps into the future. Therefore, one variable subset can present lower errors for one horizon than another variable subset, but for another horizon the situation can be inversed. Because of this, the subset that exhibits the best performance for one horizon will not necessarily be the best subset for all other horizons. To allow the comparison of subsets from a general perspective some measure of global performance must be used. In the present case we use the average performance over all forecast horizons as the measure of subset performance. In other words, the performance criterion associated to each subset will be:

$$Crit_{Sub} = \frac{1}{h} \sum_{k=1}^{h} Crit(k), \tag{5.8}$$

where $Crit(k)$ is the error criterion computed for horizon k (in the present case $\mathrm{NRMSE}(k)$ or $\mathrm{NMAE}(k)$ as defined in section 2.3) and h is the total number of forecast horizons for which forecasts are provided. When computed for the NRMSE, $Crit_{Sub}$ will be noted NRMSE_{Sub} and when computed for the NMAE, it will be noted NMAE_{Sub}.

This criterion does not privilege any forecast horizon and we choose to consider only the NRMSE_{Sub} to determine the best subsets. These, however, are both somewhat arbitrary choices and only due to the need of defining a reasonable evaluation criterion. In an operational setting, different measures could be envisaged in order to take into account specific end-user needs. In this perspective, weighted averages could for example be used to give more importance to certain forecast horizons. In the same way, other criteria such as the NMAE could be considered or even more complex error indexes that take into account several moments of the error distribution.

Another possibility could be to specifically search for the best variable subset for each horizon. However, a per-horizon search is likely to prove cumbersome and the gain in performance can be questioned when considering the complexity of defining horizon-dependent explanatory variables in an operational setting.

5.3.4 Variable subset characteristics

The aim in this section is to investigate several characteristics of the variable subsets that can prove useful to determine which subsets will be lead to the best forecasts. These characteristics will later be used to reduce the number of subsets examined by the selection methods proposed in section 5.5

Variable subset cardinality

The first variable subset characteristic that we examine is the cardinality of the subsets as it relates to the minimum and maximum average NRMSE and NMAE over all horizons. The aim of this characteristic is to determine how different input dimensions affect the performance of the model. It will allow us to determine to what extent using a higher number of explanatory variables can be beneficial to forecast performance.

Also, the possible effects of the curse of dimensionality will be revealed. As mentioned earlier, increasing the number of variables taken into account by a model usually leads to an increase in the number of parameters of the model and hence to a risk of overfitting. Another source of performance decrease can be the use of either too many redundant or irrelevant explanatory variables. If such variables are present in the initial variable set, then at some cardinality they will necessarily be included in the variable subsets and are also likely to lead to a reduction of model performance.

Reference wind farm properties

One possibility to establish the properties of a variable subset is to investigate the properties of the reference wind farms that provide the explanatory variables it contains. The idea here is to determine if some relation exists between the reference farms' properties and the forecast performance obtained using the variable subset.

Here, we propose to examine the nominal capacity of the reference wind farms and the correlation of their production to the total regional production. Examining the nominal capacity seems reasonable, given that explanatory variables provided for a subset of wind farms should provide useful data on the future production of those wind farms at least. Therefore, if the reference wind farms represent a large share of the total capacity, it can be supposed that the forecasting model will be able to correctly forecast a large share of the future production. Likewise, if the explanatory variables correspond to wind farms whose production is highly correlated to the total production, it can be supposed that these variables will provide useful information on the future regional production.

To determine the nominal capacity and correlation value that will be associated to each variable subset we propose to use the same approach for both properties: the sum of the reference wind farm properties associated to each explanatory variable subset. The reference wind farm property value that will be associated to each explanatory variable subset is defined as:

$$SubVal_i = \sum_{j \in Rsub_i} Prop_j \tag{5.9}$$

where $SubVal_i$ is the property value associated to explanatory variable subset i, $Prop_j$ is the property under consideration (nominal capacity or correlation) of reference wind farm j, $Rsub_i \subset \mathcal{A}_r$ is the subset of reference wind farms that provide explanatory variables to variable subset i and $\mathcal{A}_r$ is the set of comprising all reference wind farms.

If a strong relation between the forecast performance and the sum of nominal powers, or the sum of correlations exists, it can constitute a useful variable selection guideline. For example, a threshold based on the capacity of the wind farms could be defined so that only variables, whose reference farm capacity is above a certain level, are examined by an automatic selection procedure.

Spatial distribution of explanatory variables

As shown in chapter 3, the spatial distribution of the wind farms over the region affects the profile of the observed regional production. The spatial distribution of the explanatory variables will be examined as a criterion for variable subset selection. Most explanatory variables can be considered as providing information representative of a geographic location. Therefore, an explanatory variable subset can be considered as containing information that is more or less representative of the situation over the region. For example, if a variable subset contains wind speed forecasts for the sites of reference wind farms concentrated in a small area, the subset is likely to provide limited information on the future wind conditions for other areas of the region, and hence lead to low forecast accuracy. Conversely, if the explanatory variables in a subset are provided for locations distributed evenly over the entire region, it is possible that the subset will permit accurate forecasts.

As mentioned in chapter 3, defining a simple criterion to evaluate the geographical distribution of reference wind farms is far from trivial. Furthermore, an important question here is the spatial coverage of the region attained by the variables in a subset. To examine this aspect we propose a qualitative analysis of the geographical origin of the variables contained in the explanatory variable subsets. To this end, we will first determine the best subset, according to NRMSE_{Sub}, for each cardinality. We will then visually examine the geographic origin of the explanatory variables contained in the best subsets. In this way, it will be possible to determine the influence of the geographical origin of the variables on the observed forecast performance.

Transition stability between the best variable subsets

In addition to the above criteria we propose here to investigate if the best variable subset of cardinality $n+1$ can be derived from the best subset of cardinality n by adding a single explanatory variable to the subset of cardinality n. If this is the case, we consider that the transition between both subsets is stable. To measure the transition stability we introduce the following measure:

$$Stab(S_A, S_B) = \frac{|S_A \cap S_B|}{\min(|S_A|, |S_B|)} \tag{5.10}$$

where S_A and S_B are the subsets of explanatory variables being compared, $|S_A|$ denotes the cardinality of subset S_A, $\min$ denote the minimum function. The stability measure, as defined in Equation 5.10, varies between 0 for unstable transitions and 1 for stable transitions. This measure can be applied to the best subsets obtained for each cardinality. With this objective, the aim is to compute $Stab(S_i^*, S_{i+1}^*)$ where S_i^* is the best subsets obtained for cardinality i and S_{i+1}^* the best subset obtained for cardinality $i+1$.

Determining the stability of the transitions between the best variable subsets of successive cardinalities can be of interest in the frame of a search for the subset providing the best forecasting performance. Indeed, if the best n-variable subset is known, and the transitions can be considered to be stable, then the best subset of higher cardinality is likely to be among the subsets of cardinality $n+1$ whose transition stabilities with the subsets of cardinality n are high. Based on this hypothesis, a search heuristic can be implemented by evaluating only those subsets of cardinality $n+1$ whose transition stability with the best subset of cardinality n is above a certain threshold.

Horizon-dependent performance of the variable subsets

Up to this point we have investigated the general performance of different variable subsets without regard to the forecast horizon. However, the measure used to determine the best subsets neglects forecast horizon by considering the average performance over all horizons. We therefore propose to investigate the behaviour that can be observed across forecast horizons for the subsets determined to be the best using the NRMSE_{Sub} measure.

One aspect that can be of more interest is the cardinality of the subset that leads to the best performance for each horizon. Indeed, given the measure used to determine the best subsets and the stochastic nature of the process under consideration, it is not likely that the best on-average subset leads to the best performance for all horizons. However, comparing the cardinality of the subsets that lead to the best performance for each horizon, to that of the best on-average subset, some insight on the inherent noise content of the variables can be gained.

Given that the best on-average subset is likely to under-perform for some horizons, we will also examine the improvement that could be attained by using the best subsets for each horizon, instead of using the same subset over all horizons. From this, we will be able to quantify the performance gains that could be expected by implementing a model capable of exploiting different explanatory variable subsets depending on the forecast horizon. A final aspect we will investigate is the potential improvement that can be achieved by using the best on-average variable subset over a randomly selected subset. To this end, the performance obtained using the best on-average subset is compared to the mean performance of all subsets, for all forecast horizons. This comparison allows quantifying the performance gain due to the use of the best on-average subset compared to the performance that can be expected from the use of a random subset. In this way, the relevance of the average NRMSE over all horizons as a subset selection criterion can be verified.

5.4 Evaluation of the Impact of Variable Subset Characteristics on Forecast Accuracy

Here we present the results of the evaluation framework described in the previous section, when applied to the data from the Danish and Irish case studies. Initially we detail the variables we examine, as well as the regional models employed. We then present and analyze the results obtained for both case studies.

5.4.1 Case studies

We propose to examine two types of variables: 10 meter a.g.l. wind speed forecasts and the measured wind farm outputs; other variables such as direction or temperature will not be investigated. By considering only these variables, the number of variables falls from 143 to 22 in the Irish case and from 414 to 46 in the Danish case. Further we will examine these two variables independently. Indeed, forecast and measured variables have different impact on wind power forecast accuracy. The accuracy of the power measurements makes these variables highly explanatory of local situations and meaningful for short-term horizons. Forecast variables are less accurate but provide valuable information for medium and long-term horizons. However, the inaccuracy of these variables introduces noise in the models. If the noise in the variables is correlated, it can have highly undesirable effects on the final wind power forecast accuracy. Examining the forecast accuracy derived from measured and forecast variable subsets separately will allow a better analysis of the dependencies that exist between explanatory variables and power forecast accuracy.

The second reason which justifies the separate examination of the forecast and measured variables is that, although possibly tractable for the Irish case, examining $2^{46} - 1 \approx 7.03 \times 10^{13}$ subsets in the Danish case remains a nearly impossible task. By examining the variables separately, only 11 variables are considered at one time for the Irish case and 23 for the Danish one. This leads to the evaluation of $2^{11} - 1 = 2047$ subsets for the Irish case and $2^{23} - 1 = 8,388,607$ subsets for the Danish case.

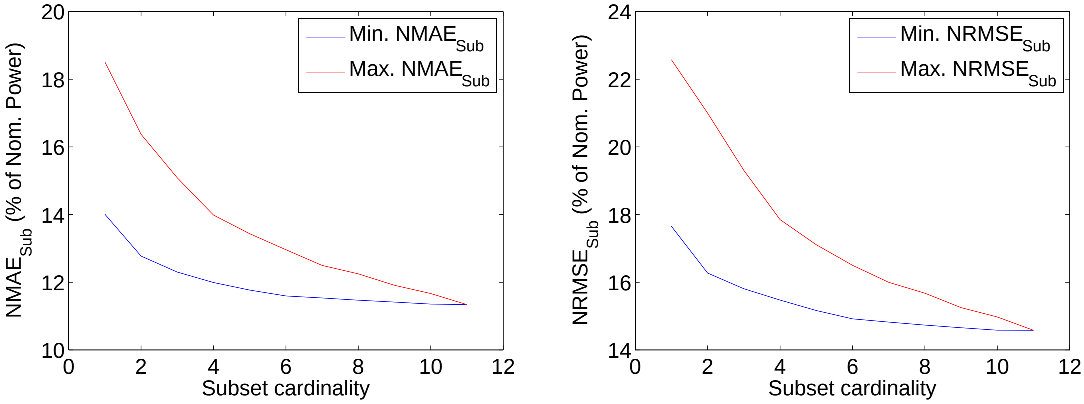
Another variable that will be considered implicitly in this study is the forecast horizon. Because it will be considered in all variable subsets, it was not considered in the computation of the number of subsets to be examined.

The forecasting model defined in Equation 5.7 will be applied here in order to evaluate the different explanatory variable subsets. The single wind farm model, which will be used as base model to examine the power measurements, will be the Persistence model. With this model only forecasts up to 12 hours ahead will be considered. Indeed, after the 12-hour horizon, the Persistence model most often under-performs models based on NWP forecasts. In the case of wind speed forecasts, the base model considered will be the RPC model presented in chapter 4. For each wind farm, the RPC model will be tuned to forecast the wind farm's future production using as sole explanatory variable the 10 meter a.g.l. wind speed forecast.

This modelling approach will be used to evaluate all variable subsets for both the Irish and Danish case studies presented in section 3.5. When considering the Irish data, the regional production corresponds to the sum of the productions of the 11 reference wind farms available for the case. In the Danish case the regional production is the power produced by all wind farms installed in the Jutland-Funen area, representing 2.2 GW of wind capacity. The datasets available for both cases are divided into a learning and a testing set. The model parameters will be estimated on the learning set and the performance of the models will be evaluated on the testing set. The division of the data into the two sets is the same as that presented in section 4.5.

5.4.2 Results for the Power Measurement Subsets

In this section we will examine the wind power measurements subset characteristics and their relation to the forecast performance obtained by the aggregation of Persistence forecasts provided for the single wind farm productions. The criteria examined are those defined in subsection 5.3.3.



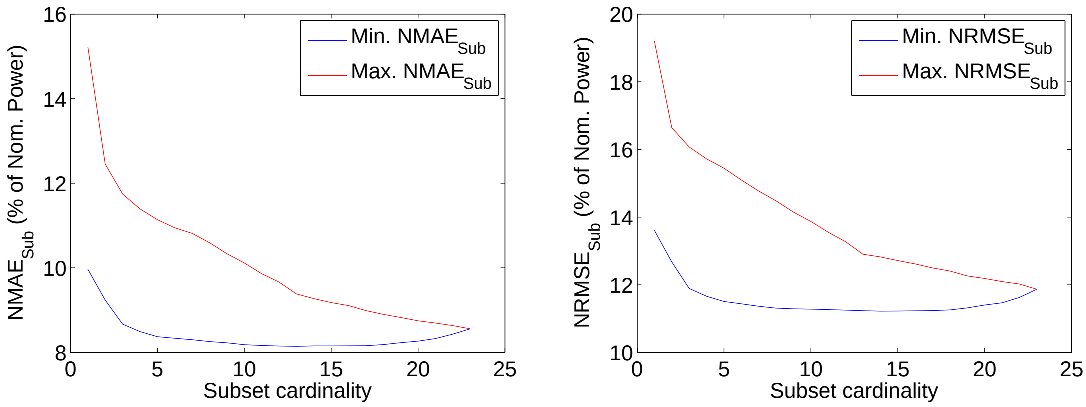
(a) Ireland NMAE (b) Ireland NRMSE

FIGURE 5.2: *Minimum and maximum* NMAE_{Sub} *and* NRMSE_{Sub} *values obtained for each subset cardinality for the Irish case study using power measurement combinations.*

Variable subset cardinality

The first aspect we will investigate is the impact of the cardinality of the variable subsets on forecast performance. In Figure 5.2 the maximum and minimum NMAE_{Sub} and NRMSE_{Sub} values obtained for the subsets of each cardinality are presented for the Irish case. For both criteria the behaviour is very similar for both the minimum and maximum values. Both the minimum and maximum values present a sharp initial decrease, which then slows for cardinalities superior to 5. In both sub-figures the minimum and maximum values converge toward the same value for cardinality 11. This convergence is expected, given that only 11 variables are considered in the initial set of explanatory variables. Only one 11-variable subset can be generated from the initial variable set and therefore for cardinality 11 a unique value is available.

Two main remarks can be formulated from these results. The first is that, as expected, considering an increasing amount of power measurements leads to a reduction of both the maximum and minimum values found for the error criteria. The second observation is that the minimal error values are obtained for the variable subset that contains all available variables. This is rather interesting given the risk of overfitting and variable redundancy mentioned earlier. In the present case, the best variable subset (over all horizons) is the one that contains the most explanatory variables. Therefore, it can be concluded that the model does not encounter parameter estimation problems and that, although the production measurements are correlated (therefore redundant) to a certain degree, this does not adversely affect the model's performance. This behaviour can be explained by the fact that the explanatory variables can be considered as noiseless, in the sense that the sum of the 11 variables equals the dependent variable. Also the regional model (see Equation 5.7) corresponds to the true model of the aggregated power and the number of observations available in the data set is sufficient for accurate model parameter estimation. Indeed, the regional power is obtained by summing the productions of the 11 wind farms. Since the dependent



(a) Denmark NMAE

(b) Denmark NRMSE

FIGURE 5.3: *Minimum and maximum* NMAE_{Sub} *and* NRMSE_{Sub} *values obtained for each subset cardinality for the Danish case study using power measurement combinations.*

variable is the sum of the explanatory variables the aggregation model, which is linear, can easily be parameterized, at least for horizon zero. Finally, in the Irish case study the number of observations in the training set is sufficiently large with respect to the total number of parameters to be estimated in the case where all variables are considered.

In Figure 5.3 the maximum and minimum NMAE_{Sub} and NRMSE_{Sub} values obtained for each subset cardinality are presented for the Danish case. From a general standpoint, these results are similar to those found in the Irish case. For both cases the maximum and minimum NMAE_{Sub} and NRMSE_{Sub} values present a sharp initial decrease, which then becomes less pronounced. However, in the Danish case the minimum values for both criteria present an increasing trend after cardinality 12. For this same cardinality, an inflection point is visible in the maximum error criteria curves; the rate of decrease of the maximum values changes at cardinality 12. The increase in minimum values observed for higher cardinalities is symptomatic of the curse of dimensionality.

More specifically, the behaviour of the minimum NMAE_{Sub} and NRMSE_{Sub} values in the Danish case is typical of the results which can be expected when a prediction model is based on an increasing number of explanatory variables. Although the explanatory variables are noiseless and the model corresponds to the true relation between the explanatory and dependent variable, increasing the number of explanatory variables leads to a saturation in the reduction of the error and even to a decrease of model performance when too many variables are considered. Several factors can help explain this behaviour. The factor which can be linked to the curse of dimensionality is that the model can not adequately estimate its parameters when more than roughly 12 variables are considered given the number of observations in the learning set, in the present case 543 observations. When too many variables are considered, the model parameters are too numerous and their estimation is no longer statistically significant given the size of the learning dataset. In other words, given the dimension of the input space, the observations in the learning set are too few to adequately describe the relation between the explanatory and dependent variable.

Another factor that can be considered here that in the present case only the production measurements of 23 wind farms, out of the tens installed in the Jutland-Funen peninsula, are considered as explanatory variables. The dependent variable is the aggregated production of all wind farms installed within the region, not the aggregated production of the 23 wind farms. The available information on the process under consideration is incomplete. Under these circumstances any model will necessarily lead to some error.

This "incomplete" information situation in the Danish case study can be compared to the "perfect" information of the Irish case study. In the Irish case all the information on the regional production was available; therefore the total production could be perfectly calculated from the available wind farm production measurements. It can therefore be thought that having measurements for all wind farms would do away with the error increase after a given cardinality. This however is not certain, as will be shown later in this section.

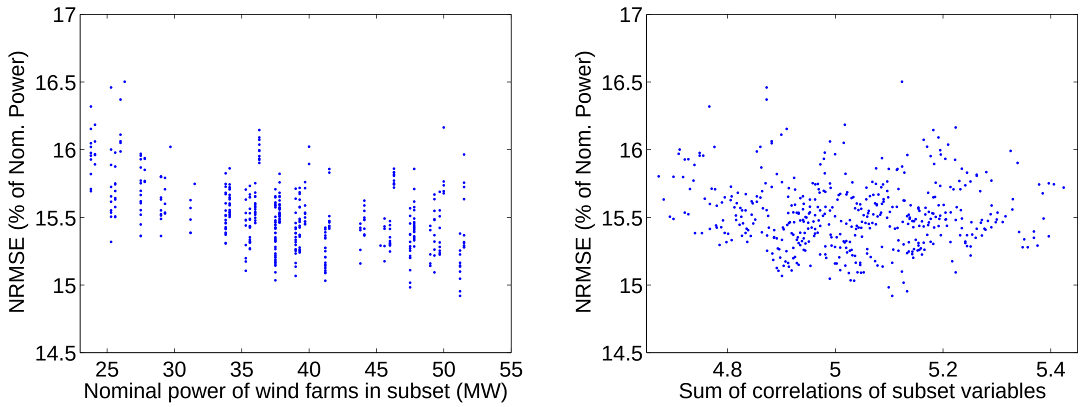
From these results it appears that when considering measured variables, or forecasts derived from measured variables (in this case simple Persistence forecasts) many variables can be taken into account without adversely affecting forecast performance. However, in upscaling cases where "incomplete" information is available, the subsets that lead to the best performance are not necessarily those that incorporate all available explanatory variables. In such cases it is then necessary to examine which explanatory variables lead to the best forecasts. The aim is of course to identify the characteristics of these variables so that a selection process can be determined.

Reference wind farm properties

The aim here is to determine if reference wind farm properties can help determine the best explanatory variable subset. As mentioned earlier, the least squares regression, on which the regional model is based, minimizes the sum of square errors. Therefore, we will limit the examination to the relation between the reference wind farm properties (nominal power and correlation of production to total production) and the NRMSE_{Sub}.

In Figure 5.4 the relation between nominal power and forecast error as well the relation between correlation and forecast error are examined for 6-variable subsets in the Irish case. In Figure 5.4(a), the NRMSE_{Sub} obtained for each variable subset is plotted against the sum of the nominal powers of the wind farms whose production measurements are combined. In Figure 5.4(b) for each subset the correlation coefficients between the explanatory variables and the total production are summed to provide a measure of the correlation between a subset's variables and the dependent variable. The summed correlations of each subset are plotted against the NRMSE_{Sub}.

From the results presented in these figures it is clear that nominal power alone is not sufficient to determine the subset that leads to the lowest error forecasts. Although the error does tend to decrease as the nominal power represented by the variables in each subset increases, the best forecasts are not achieved for the subset with the highest share of nominal power. Further, it is clear from the vertical alignment of several points in the scatter plot



(a) Sum of nominal powers

(b) Sum of correlations

FIGURE 5.4: *Sum of subset nominal powers and sum of subset correlations between wind farm productions and total production, versus* NRMSE_{Sub}*. The values presented here are for the Irish case, and 6-variable subsets. The variables under consideration are measured wind farm productions.*

that subsets whose nominal power is identical can lead to widely varying levels of forecast error.

Hence, when considering wind farm power measurements, choosing the subsets that represent the highest share of the total nominal power will not necessarily lead to the lowest possible error but it can be noted that it does improve the likelihood of choosing a subset that leads to reasonably accurate forecasts.

The lack of relation between the sum of the correlations between the dependent variable and the explanatory variables belonging to a subsets and the subset's forecast error is more pronounced than that observed between the subset's nominal power and the forecast error. When considering correlation, the best forecasts are obtained by subsets whose cumulated correlations are neither the highest nor the lowest, but rather seem to be located in the middle of possible correlation values. Also, the dispersion of forecast errors is rather high even for subsets whose correlation values are close to those of the subsets that lead to the best performance. This lack of relation between the correlations and forecast performance can be explained by that fact that subsets whose aggregated variable correlations are high comprise variables that are all highly correlated with the dependent variable. However, this high correlation with the dependent variable can also mean high correlations between the variables of the subset and thus a high level of redundancy between variables. The opposite effect can explain the high errors observed for subsets with low correlation variables. The variables are not likely to be redundant but their relevance to the dependent variable is low; they do not provide the forecast model with as much useful information as more highly correlated variables.

FIGURE 5.5: *Locations of the wind farms whose production measurements belong to the best variable subsets of each cardinality. The results are for the case of Ireland.*

Spatial distribution of explanatory variables

The two previous parameters only partially explain the predictive power of different variable subsets. Another characteristic of the variables that can be investigated is their geographical origin and distribution. The results presented in this section are very similar for both case studies. We therefore only present the results for the Irish case, the results for the Danish case can be found in Appendix C.

In Figure 5.5 the locations of the reference wind farms that provide the production measurements belonging to the best variable subsets of cardinalities 1 to 6 are presented for the Irish case. For cardinality 1, the wind farm that provides the power measurements that lead to the best forecasts is wind farm 3. This location of this wind farm, in relation to the other wind farms in the region, is rather central. Most wind farms are located in northwestern Ireland and wind farm 3 is centrally located with respect to nine other wind farms.

The subset that leads to the best forecasts for cardinality 2 comprises measurements from wind farms 3 and 8. When compared to the best 1-variable subset, we notice that wind farm 3 remains and that wind farm 8 is also considered. Wind farm 8 is located closer to the northern farms. We can also mention that wind farm 8 represents more than 20% of the installed capacity in the region and its production is the most highly correlated to the total production: with a correlation coefficient of 0.95.

TABLE 5.2: *Subsets presenting the best average NRMSE over all horizons for the first 6 cardinalities, for the case of Ireland, considering wind power measurements.*

Cardinality	Wind farm indexes	Cardinality	Wind farm indexes
1	3	4	1 2 3 6
2	3 8	5	1 2 3 5 6
3	1 2 3	6	1 2 3 5 6 8

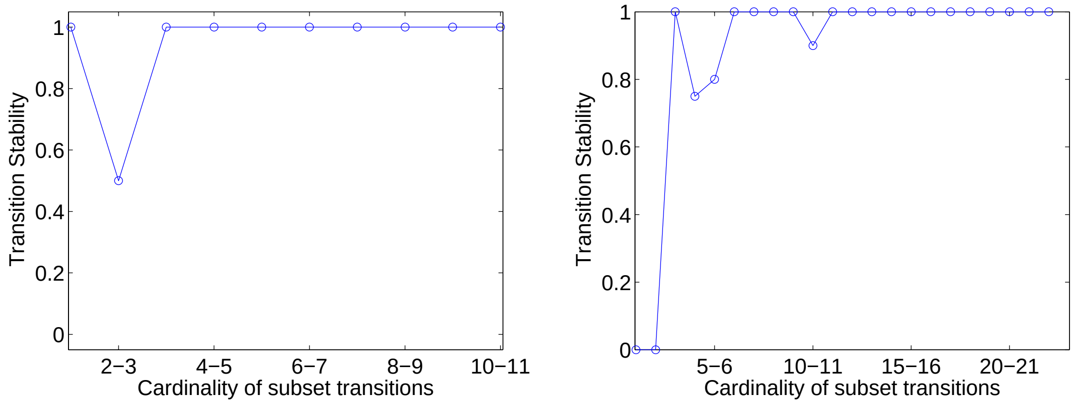
The best subset of cardinality 3 includes power measurements from wind farms 1, 2, and 3. For this subset we can notice that wind farm 3 is again included while wind farm 8 is no longer considered. For this cardinality, wind farm 2, which is located in a more northerly location, "replaces" wind farm 8. Wind farms 2 and 8 have the same nominal power, but the correlation between the production of wind farm 2 and total production (0.93) is lower than that of wind farm 8. Wind farm 1 is located to the west of the two other wind farms and its production is likely representative of the productions of the most westerly located wind farms.

When examining the best subset of cardinality 4, it is interesting to notice that it only differs from that of cardinality 3 by the "addition" of measurement from wind farm 6, the only wind farm located in the south of the region. This fact is of interest since wind farm 6 represents less than 7% of the total nominal power and given its location, its production presents the lowest correlation to the total production (0.62) and to the other wind farm's productions. However, this variable subset offers a good coverage of the region with wind power measurements for all the different wind regimes that can occur over the region under consideration.

This is somewhat confirmed by the best subsets for cardinalities 5 and 6. The reference wind farms associated to the best subset of cardinality 4 are also associated to the best subsets of cardinalities 5 and 6, and it can be mentioned that the best subset for cardinality 6 incorporates all the variables present in the best subset of cardinality 5. For cardinality 5 the extra variable considered, with respect to the subset of cardinality 4, is again provided by a wind farm located in an "extremity" of the region, being the northern most wind farm.

For cardinality 6, the added variable is provided wind farm 8 which had already been considered for cardinality 2. It is likely that the subset of cardinality 5 provides a good enough coverage of the region, and that for cardinalities beyond cardinality 5, the variables that will be incorporated into the best subset will be those come from wind farms that have the highest share of the total nominal power, combined to productions that present the highest correlation to the total production.

From these results, it is clear that the geographical location of the wind farms that provide the power measurements plays a role in the forecast performance that can be achieved from their combination. The location of the wind farms is relevant in that it is linked to differing wind regimes. From these results it also appears that there is certain stability in the evolution of the best subsets for increasing cardinalities as shown in Table 5.2.



(a) Irish case

(b) Danish case

FIGURE 5.6: *Transition stability between best production measurement subsets, of consecutive cardinality, for the Irish and Danish case studies.*

Transition stability between the best variable subsets

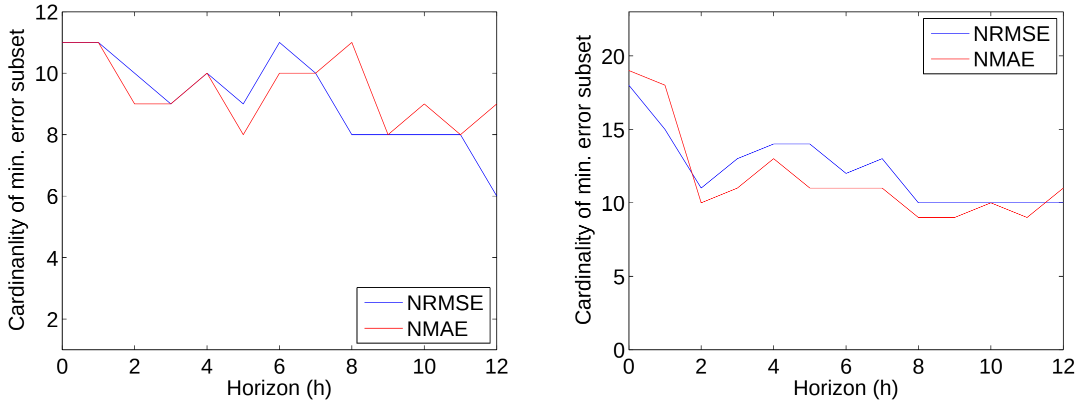
In the Irish case presented above, the best subset for a cardinality can in most cases be derived from the best subset of lower cardinality by adding one wind farm. For example the best subset of cardinality 4 comprises wind farms: 1-2-3-6. This subset can be derived from the best subset of cardinality 3 (wind farms: 1-2-3) by adding wind farm 6. From this remark it can be of interest to examine the stability of the transitions between the best subsets of one cardinality to that of the next highest cardinality.

In Figure 5.6 the stability of transitions between successive subsets cardinalities, considering only the subsets that lead to the lowest average NRMSE for each cardinality are presented for both case studies. The subsets of each cardinality for which the lowest average NRMSE is obtained can be found in Appendix C. In this figure it is clear that for both cases the transitions are rather stable once sufficiently high cardinalities are considered. In the Irish case the only unstable transition is the transition between the subset of cardinality 2 and the subset of cardinality 3. In the Danish case the transitions are more unstable for low cardinalities. The transitions between cardinalities 1 and 2, and between cardinalities 2 and 3 are totally unstable. This means that best subsets for cardinalities 1, 2, and 3 share no common variables. For higher cardinalities the transitions stabilize somewhat, which means that the best subsets for higher consecutive cardinalities comprise a large share of common variables.

Horizon dependent performance of the variable subsets

Up to this point we have investigated the general performance of different variable subsets without regard to the forecast horizon. In the following paragraphs we investigate the horizon-dependent performance of the variable subsets.

In Figure 5.7 the cardinalities of the subsets leading to the lowest NRMSE and NMAE



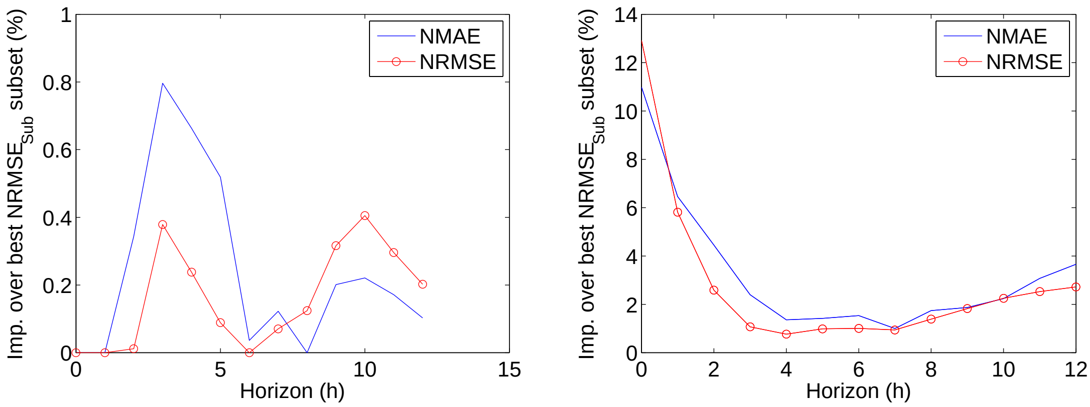
(a) Irish case (b) Danish case

FIGURE 5.7: *Cardinality of the measured power forecast subsets leading to the lowest NRMSE and NMAE values for each forecast horizon. It must be noted that the subsets that lead to the lowest NRMSE and lowest NMAE are not necessarily the same for each horizon.*

values for each forecast horizon are presented for the Irish and Danish cases. The first striking aspect of these figures is that the cardinality of the best subsets for each horizon is not constant. Indeed, the cardinalities in both cases vary between 8 and 18 variables. This variability can be explained by slight differences in NRMSE and NMAE values for subsets of different cardinalities, which might not necessarily be very significant from one horizon to the next. However, an interesting aspect that is present in both cases is the progressive reduction in cardinality of the best variable subsets as the forecast horizon increases. In both cases, the cardinality of the subset leading to best forecasts for horizon 0 is the maximal, or near maximal, cardinality, whereas for horizon 12, this value falls to between 6 and 11 depending on the error criteria and the case study. This reduction in the cardinality of the best subsets, as the horizon increases, can be explained by the increasing uncertainty, or noise, in the relation between the Persistence forecasts for the single wind farm productions and the total production. The fact that higher amounts of noise are present for higher forecast horizons renders parameter estimation more difficult.

We can also notice that in the Irish case the best variable subsets for higher horizons have relatively low cardinalities, 6 variables for horizon 12, whereas in the Danish case, the cardinalities are higher: 10 variables for horizon 12. The difference in cardinality reduction between the Irish and Danish cases can also be due to the fact that, in the Irish case, the total regional production is provided by 11, geographically concentrated, wind farms, while in the Danish case, the regional production is provided by tens of wind farms, representing 2.2 GW of installed capacity and more evenly distributed across the region. Because of this, the Danish regional production presents higher autocorrelation than the Irish one, see subsection 3.6.1. Therefore, the persistence forecasts of the individual wind farm productions in the Danish case provide more useful information to the regional model even for higher horizons than those of computed in the Irish case.

Another interesting aspect that is apparent from these results is that the cardinality of

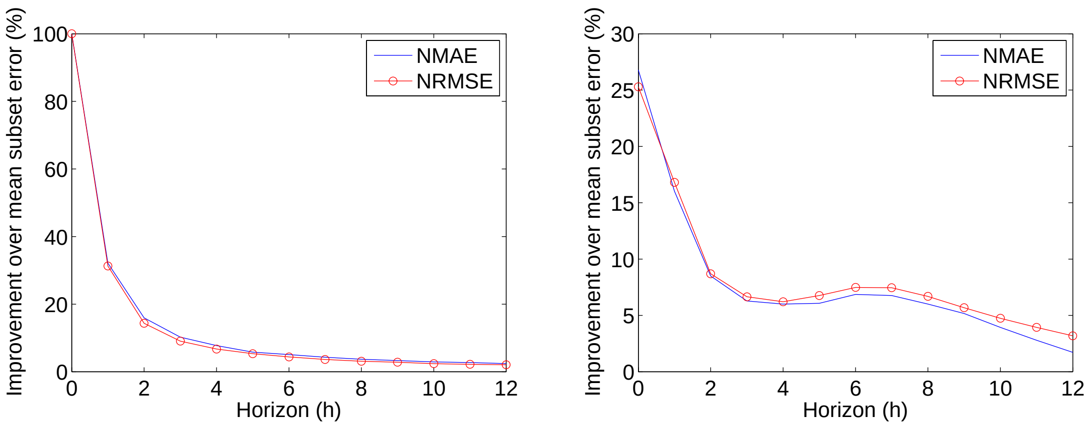


(a) Irish case (b) Danish case

FIGURE 5.8: *Improvement in forecasts error of the best subsets per horizon over the best subset in terms of* NRMSE_{Sub}*. The results are presented for the Irish and Danish case, considering persistence forecast combinations.*

the subset that leads to the best average error criterion over all horizons is not the same as the cardinality of the best subset for each horizon. This is evident in the Irish case where the best subset in terms of NRMSE_{Sub} was the one comprising all 11 variables. This subset is however only the best subset for three horizons out of the twelve examined here. The same is true for the Danish case where the cardinality of the subset leading to the lowest NRMSE_{Sub} has a cardinality of 14, whereas the best subsets cardinality for each horizon varies between 9 and 18. For the Danish case, the best NRMSE_{Sub} cardinality is only the best cardinality for two horizons. The fact that the subsets which lead to the best NRMSE over all horizons differ from those that lead to the best performance for each horizon can cast some doubt on the validity of the average error criteria as a means of selecting the best variable subset. Although this might be exact in absolute terms, in a practical context where models are installed for on-line operation, the added complexity (understand cost) of changing input variables for different horizons, might not be justified given the possibly low improvement in forecast performance.

In Figure 5.8 the improvement in terms of NRMSE and NMAE of the best subsets per horizon over the subset having the best NRMSE_{Sub} are presented for the Irish and Danish case. Clearly the difference in performance is important between the two case studies. In the Irish case, the improvement is below 0.5% for the NRMSE for all horizons. For this case, considering the best subsets per horizon does not lead to a great improvement of the forecast error. The situation for the Danish case is obviously different with improvement values up to 13% of NRMSE for horizon 0. Another interesting fact that can be noticed for this case is that when the best horizon subsets are of the same cardinality as the best on-average subset, the improvement is positive. This means that for the Danish case, the best subset on-average does not lead to the best possible performance for any of the forecast horizons. We can nonetheless remark that when the best per-horizon subset's cardinalities are close to that of the best on-average subset, the improvement is lesser than when the difference in



(a) Irish case (b) Danish case

FIGURE 5.9: *Improvement in terms of the NRMSE and the NMAE of the best on-average subset over the mean forecast error computed over all possible subsets. The results are presented for the Irish and Danish case, considering persistence forecast combinations.*

cardinality is more important. Clearly, there is a relation between differences in the number of variables in the subsets and the improvement that is observed. For this case, considering the best on-average subset for all horizons leads to a deterioration of forecast performance. The suitability of using the average NRMSE over all horizons to determine the best variable subset can be questioned in this instance.

The difference in performance of the best on-average subset for both case studies when compared to the best possible performance can possibly be due to both the dimension of the case studies and the fact that in the Irish case the available variables provide "complete" information on the dependent variable while the information provided by the available variables in the Danish case is "incomplete".

A final aspect we will investigate here is the potential improvement that can be achieved by using the best on-average variable subset over a randomly selected subset. From this point of view, the NRMSE and NMAE obtained with the best on-average subset for each horizon are compared to the expected values of NRMSE and NMAE obtained by using random subsets. In other words, the error criteria are compared to the average values obtained with all possible subsets, for each forecast horizon.

The results of this comparison are shown in Figure 5.9. In both cases there is an evident improvement[2] for all forecast horizons. Also, in both cases, the improvement decreases from rather high values for the first horizons to relatively low improvements for the long-term horizons (long-term here, being 8-12 hours ahead). The shape of the improvement curve can be linked to the nature of the single wind farm forecasts aggregated by the regional forecasting model. Indeed, in the present case, the single wind farm forecasts are persistence forecasts which usually only provide "competitive" forecast for horizons between 0 and 6 hours ahead. The "noise" present in the explanatory data increases very

[2]Here the improvement is computed according to Equation 2.15.

rapidly, which leads to similar performance for most subsets for higher forecasts horizons, leaving little room for improvement of one subset over another. Therefore, the best on-average subset does not outperform other subsets as much for higher horizons. This can be illustrated by the fact that in the Irish case the best subset, which comprises all variables, presents 100% improvement for horizon 0, while for horizon 12, the improvement over a randomly selected subset can be expected to be in the order of 2% to 3%.

Once again, there is a significant difference between both case studies. In the Irish case where "complete" information is available the improvement for short-term horizons is very important, whereas for the Danish case, the short-term improvement is much lesser. This can possibly be due to difference in "completeness" of available information between both case studies.

The main conclusion that can be formulated from this comparison, is that, when considering the aggregation of persistence forecasts, although the best on-average subsets might be "sub-optimal" for some or all forecast horizons, considering it instead of some randomly chosen subset will likely lead to better than average results for all horizons. It is then probably worthwhile to determine a method that can allow the determination of the best on-average subset, or what would be even better, given the difference between the best on-average subset performance and the best per-horizon subset performance, to determine the best variable subset for each horizon

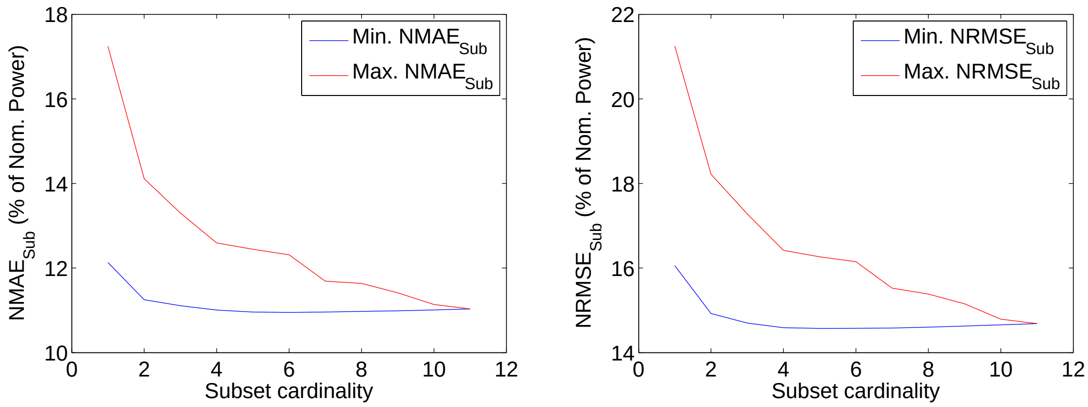
From a more general perspective some preliminary conclusions can be formulated on the selection of measured variables. As we have seen in this section, the benefit of selecting input variables is greater for horizon 0. For this horizon no time-dependent extrapolation is considered, the only aspect being considered is the ability of determining the aggregated regional production from a subset of its aggregates. It is clear that considering many, or all, production variables can be recommended as a rules of thumb approach. This however can be questioned for cases similar to the Danish one, where some "geographical" (not temporal) extrapolation is performed, and where the number of available explanatory variables is large with respect to the number of observations in the dataset. For such cases, adequately selecting a subset of the available explanatory variables is likely to often yield the "optimal" forecasts.

5.4.3 Results for the Wind Speed Forecast Subsets

In this subsection the aim is to examine the use of different subsets of wind speed forecasts. Here the approach is the same as that presented in the previous section, where forecasts based on production measurements were combined to produce regional forecasts.

Variable subset cardinality

The first results we examine are the minimum and maximum NRMSE_{Sub} and NMAE_{Sub} values over all horizons (see Equation 5.8) for each subset cardinality for both the Irish and Danish case studies. In Figure 5.10, the results are presented for the Irish data. The general



(a) Ireland NMAE

(b) Ireland NRMSE

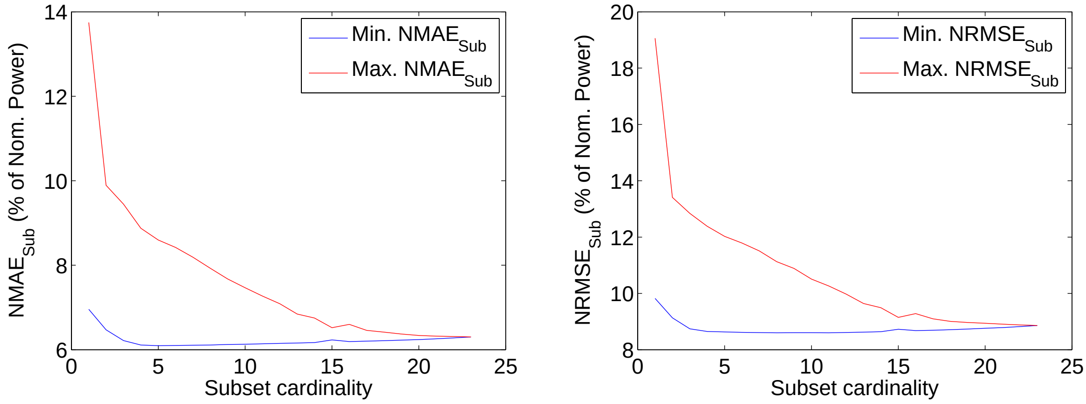
FIGURE 5.10: *Minimum and maximum average* NRMSE_{Sub} *and* NMAE_{Sub} *values obtained for each subset cardinality for the Irish case study using wind speed based power forecast combinations.*

shape of the curves is similar to those found when examining the power measurement forecasts. For low cardinalities the error is high and both the maximum and minimum values present a decreasing trend for the first cardinalities.

It can however be noticed that the NRMSE_{Sub} reaches a minimum for cardinality 5 while the NMAE_{Sub} reaches a minimum for cardinality 6. For higher cardinalities, the minimum average error values increase steadily. In this case the behaviour is similar to that observed for the power measurement combinations in Danish case. As a matter-of-fact the general shape of the curves more closely resembles the one observed for the power measurement based combinations in Danish case than the Irish case.

In the present case there is a clear difference with the power measurement case presented in the previous section. In the previous case, the explanatory variables presented no noise, at least for forecast horizon 0. In the present case, even for low forecast horizons, the explanatory variables can lead to single wind farm forecasts presenting large errors. This is similar to the Danish case presented earlier where noise could be linked to the incompleteness of the available information. In the present case the noise is of course due to the forecast errors that result from the NWP forecast errors and the modelling error of the RPC model. Again, the fact that noise is present in the explanatory variables leads to a saturation point after which, increasing the number of variables has a negative effect on the generalization capability of the regional forecasting model.

In Figure 5.11 the same analysis is presented for the Danish case, and gives similar results. In this case the lowest NRMSE_{Sub} is obtained for a subset of cardinality 8 and the lowest NMAE is obtained for a subset of cardinality 5. These cardinalities are lower than those found for the Danish case when considering the forecasts based on power measurements. This again can be linked to the superior amount of noise present in the explanatory variables. Again, for higher subset cardinalities the minimum NRMSE_{Sub} and NMAE_{Sub} increase steadily with subset cardinality.



(a) Denmark NMAE (b) Denmark NRMSE

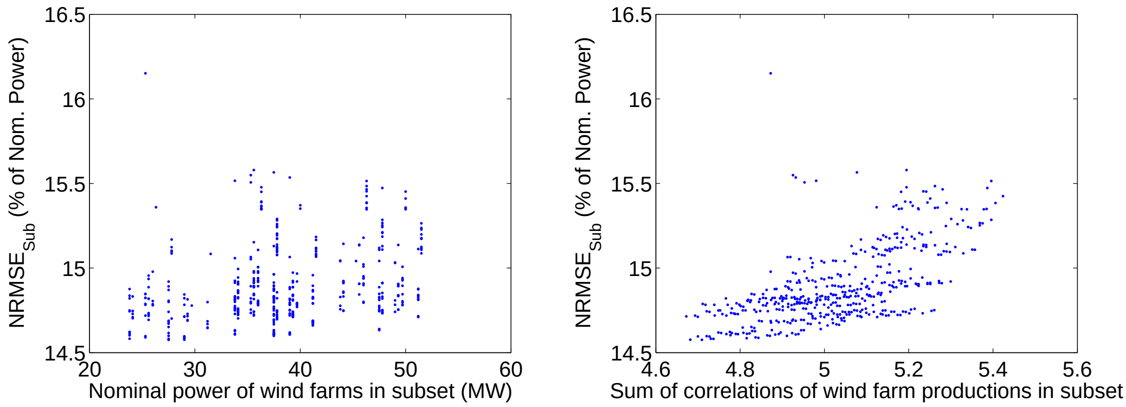
FIGURE 5.11: *Minimum and maximum* NRMSE_{Sub} *and* NMAE_{Sub} *values obtained for each explanatory variable subset cardinality for the Danish case study, using wind speed forecasts as explanatory variables.*

An interesting feature that is apparent in these results is the very narrow difference between the minimum and maximum values found for subsets whose cardinality is superior to 12. In all the cases considered up to this point, the difference between the minimum and maximum average error criteria were rather wide and the reduction of the difference with subset cardinality was rather slow. In the present case, the difference is quite slight; the difference between the minimum and maximum is less than 1% of nominal Power for the NRMSE_{Sub} results. These results therefore imply that for large subsets, the difference between the best subset and a randomly selected subset will be very small, and its statistical significance can be questioned.

Reference wind farm properties

The results found for both case studies seem to confirm the hypothesis that considering too many noisy and correlated explanatory variables can hamper the performance of statistical forecasting models by impeding adequate model parameter estimation. As with the power measurement data, the question that comes to mind is to know if the wind farms, associated to the explanatory variables that belong to the best on-average subset, possess some specific characteristic.

In Figure 5.12 the relation between the sum of nominal powers of the reference wind farms associated to a variable subset and the forecast error, as well as the relation between the correlations of the reference wind farm productions to the regional production and the forecast error, are presented for the 6-variable subsets of the Irish case. As shown in the previous section, the relation between wind farm nominal power and forecast error is tenuous if it exists at all. With respect to the sum of correlations, there appears to be a negative relation between it and the forecast error. This is somewhat surprising, since it implies that the subsets leading to the lowest errors contain explanatory variables from reference wind



(a) Sum of nominal powers

(b) Sum of correlations

FIGURE 5.12: *Sum of subset nominal powers and sum of subset correlations between wind farm productions and total production, versus* NRMSE_{Sub}*. The values presented here are for the Irish case, and 6-variable subsets. The variables under consideration are wind speed forecasts.*

farms whose production is not highly correlated to the total production. This can be explained by the fact that only possessing variables from farms whose production is highly correlated to the total production does not provide the regional forecasting model with adequate information on the meteorological situation over the entire region. Indeed, if the productions are highly correlated to the regional production, they will also be highly correlated amongst themselves; the reference wind farms are then likely to be located near each other. Therefore, the explanatory variables they provide lack sufficient geographical coverage. This tends to confirm the results of the previous section where the geographical location of the reference wind farms associated to the best subsets were shown not to be random.

Spatial distribution of explanatory variables

The results for the geographical dispersion of the variable are presented in Figure 5.13 for the Irish case for the first 6 cardinalities. The best on-average subsets for all cardinalities, for the Irish and Danish cases can be found in Appendix C, along with a figure giving the geographical location of the wind farms related to the best on-average subsets for the Danish case. For both the Irish and Danish cases the conclusions concerning the geographical distribution of the reference wind farms are the same.

For the Irish case, the variable leading to the best regional forecast is provided by wind farm 2. This wind farm is not as centrally located as wind farm 3 which provided the power measurement that lead to the best results in the previous study. It is however located near farms 4, 5, 7, and 8. These farms, including wind farm 2, represent more than 64% of the installed capacity in the region. Furthermore, the correlation of the production of wind farm 2 with the regional production is high and the correlation of its production with that of the farms located near to it is also high. Another important property is that the single

FIGURE 5.13: *Locations of the wind farms whose wind speed forecasts belong to the best variable subsets of each cardinality. The results are for the case of Ireland.*

TABLE 5.3: *Comparison of the best wind power measurements subsets and best wind speed forecasts subsets, of cardinalities 1 to 6, for the case of Ireland.*

Cardinality	Wind farm indexes	
	Wind power measurements	Wind speed forecasts
1	3	2
2	3 8	7 9
3	1 2 3	4 5 9
4	1 2 3 6	4 5 6 9
5	1 2 3 5 6	4 5 6 9 10
6	1 2 3 5 6 8	4 5 6 7 9 10

wind farm forecasts derived from the NWP forecasts for this farm present very low errors when compared to the single wind farm forecast errors observed for other farms. These properties can explain why the wind speed forecasts associated to wind farm 2 lead to the lowest regional forecast error when considering a single explanatory variable.

For cardinality 2, the wind farms that lead to the lowest on-average NRMSE are wind farms 7 and 9. The location of these wind farms is significantly different from the best two-variable subset found for the power measurement case. Indeed, for the power measurement case, wind farms 3 and 8 are both somewhat centrally located, and near to each other. Since wind farms 7 and 9 in the present case are clearly distant, the variables they provide can be said to "cover" a much larger part of the region under consideration.

With regard to the three-variable subset, wind farms 7 and 9 are also associated to it, along with wind farm 4. We can remark that wind farm 4 is close to wind farm 2, which was the "new" wind farm in the best three-variable subset found in the power measurement case. The "addition" of wind farm 4, located near wind farms 2 and 8, which represent more than 40% of the installed capacity, probably provides the forecasting model with finer information on the future production of these two farms, in particular, and of the farms located in the northernmost area of the island.

The best subset found for cardinality 4 is the same as that found for cardinality 3 with the addition of wind farm 6. This is very similar to the four-variable subset found in the power measurement case, where the power forecasts for wind farm 6 were also "added" to the best three-variable subset. The transition between the subsets of cardinality 3 and 4, points to a progressive coverage of the region.

The subset of cardinality 5 contains all the variables of the four-variable subset, with the addition of the variable from wind farm 10. Wind farm 10 is located in a central location with respect to all wind farms, and particularly close to wind farms 3 and 11. It can be noticed that wind farm 3, 10, and 11 form a small cluster in a central position. Up to cardinality 5, the regional forecasting model did not have "specific" information on these three wind farms.

The six-variable subset is very similar to that found for cardinality 5, the only difference being the variable from wind farm 7. Here again, the variable that is "added" in comparison to the five-variable subset, is related to a wind farm located in the northern most area of the region. It can also be noticed that wind farm 7 is one of the farms nearest the north coast, and that wind farms 4 and 5, which also belong to this subset, are also located near the north coast. The presence of many variables associated to northern farms can possibly be due to particularities in the NWP forecasts provided for these locations.

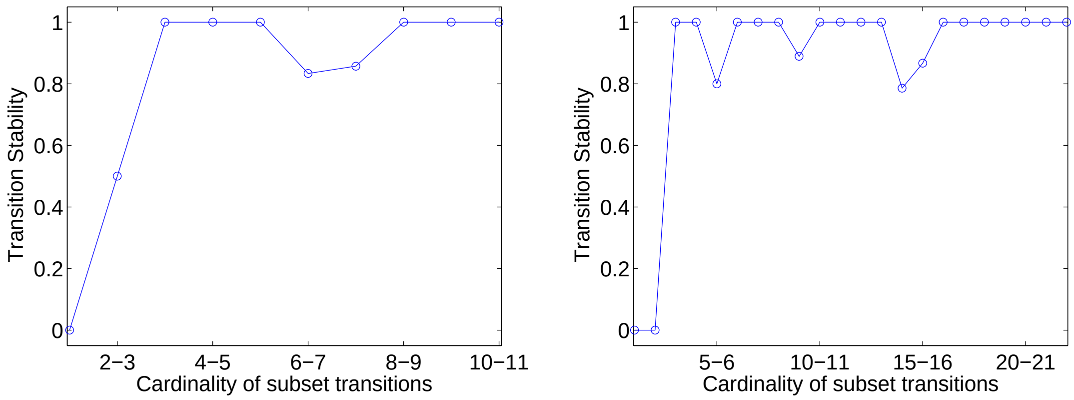
From these results it is apparent that the locations which provide the variables that lead to the best performance for different explanatory variable subsets are not random and that they are certainly linked to differing meteorological conditions over the region. In the previous case, where the explanatory variables were derived from power measurements, the correlation between wind farm productions, and the nominal powers of the wind farms played a more apparent role. In the present case the explanatory variable are based on wind speed forecasts. The forecast errors are therefore more closely linked to the meteorological representativeness of the NWP variable locations than to the nominal power of the wind farms or the correlation of their production to the total production. Another factor that can help explain the higher geographical dispersion of the selected variables, especially for lower cardinalities, is the fact that, for higher horizons, the NWP forecast error correlation between grid points increases [24]. Since correlated explanatory variables are prone to be redundant, selecting farms that are far apart can help lower the correlation of the errors and hence reduce forecast error by statistical smoothing.

Transition stability between the best variable subsets

As shown in the results presented in the above paragraphs, the transitions between the best on-average subsets of different cardinalities seem rather stable. Here we investigate the transition stabilities between all cardinalities as described in the previous section. We apply the stability measure presented in Equation 5.10 to the best on-average NRMSE variable subsets of successive cardinalities, for both the Irish and Danish case. The aim is the same as in the case of the power measurement combination case, that is, to investigate if the best subset for cardinality $n+1$ can be derived from the best subset of cardinality n.

In Figure 5.14, the transition stabilities between the best subsets (in terms of NRMSE_{Sub}) for successive cardinalities are presented for both cases. From a general point of view, the stability transition curves for both cases present similar shapes with high instability for the transitions between the first few subset cardinalities, followed by very stable transitions for some cardinalities and then a few slightly unstable transitions for higher cardinalities.

In both cases it can be noticed that the first two transitions are highly unstable. This instability can be linked to the geographical location of the wind farms whose forecasts belong to the best subsets. As we have seen for low cardinalities, the best subsets are those that provide good “geographical coverage” of the region. Therefore, when only variable from a few reference farms are considered, the best coverage is attained with widely different subsets. Once a certain number of variables are considered, a sufficiently representative coverage is probably attained. Additional variables only refine the coverage but do not allow major coverage improvements with respect to the best lower-cardinality subsets.



(a) Irish case (b) Danish case

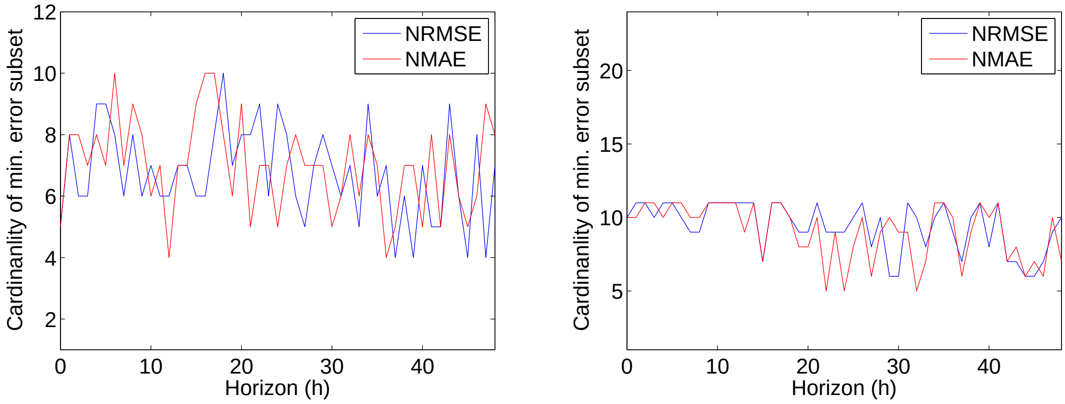
FIGURE 5.14: *Transition stability between best wind speed forecast subsets, of consecutive cardinality, for the Irish and Danish case studies.*

Another interesting aspect that appears in the transition stability curves is the low, but none the less present, transition instability that can appear for cardinalities above 5. In both cases some instability is present for some subset transitions. However, it must be noted that the transitions leading to the overall best subsets for both cases are stable and the transition from the best overall subset to the one of higher cardinality is also stable. This phenomenon can possibly be linked to the low difference in forecast performance between the best subsets for each cardinality and the subset leading to the best overall performance.

When comparing the transition stability results found for the wind speed forecasts to those found for the power measurements, it appears that there is higher instability in the case of the wind speed forecast subsets. The higher instability is due to the higher noise content present in the wind speed based forecasts that are combined. This is comparable to the difference in stability observed in the power measurement case between the stable Irish transitions and the more unstable Danish transitions. Here the difference is also marked between the Irish and Danish results, the Irish transitions being much more stable than the Danish ones. This can possibly be explained by the fact that, in the Irish case, wind speed forecasts are available for all the wind farms considered in the region, whereas in the Danish case, they are only available for a small fraction of the wind farms whose aggregated production is being forecast.

Horizon-dependent performance of the variable subsets

As we have seen the transitions between the best on-average subsets for each cardinality are rather stable in the present case although the explanatory variables are noisier than in the power measurement case. To further examine the effect of the higher noise content, we investigate the cardinality of the subsets that lead to the best performance for each forecast horizon. The approach is the same as the one used in the power measurement case described in subsection 5.4.2.



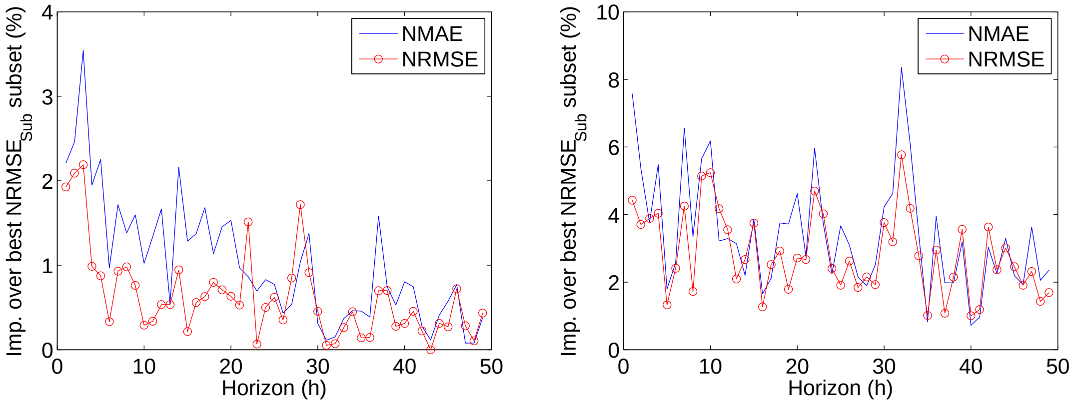
(a) Irish case (b) Danish case

FIGURE 5.15: *Cardinality of the wind speed forecast subsets leading to the lowest NRMSE and NMAE values for each forecast horizon. It must be noted that the subsets that lead to the lowest NRMSE and lowest NMAE are not necessarily the same for each horizon.*

In Figure 5.15 the cardinalities of the subsets leading to the lowest NRMSE and NMAE for each forecast horizon are presented. Clearly the cardinalities of the subsets that lead to the best results both for the NMAE and NRMSE are rather variable. Although the variability appears to be higher in the Irish case, it must be noted that the scale of the graphic is smaller than the one used in the Danish case. Actually the variability is to a certain extent comparable, the amplitude of variation reaching 6 in both cases.

The Irish case does however present a higher frequency of variation with virtually no constant cardinalities from one horizon to the next. In the Danish case, the cardinalities are more stable, with several successive horizons presenting best subsets of equal cardinality. In both cases the variability appears to increase with forecast horizon. This can be due to the higher noise present in the explanatory variables for higher horizons, resulting from the increase in forecast error with forecast horizon that is natural in wind speed forecasts.

When compared to the subset cardinalities obtained for the power measurement subsets, the results present much higher variability. Also, in the previous case, the cardinality very clearly decreased with higher forecast horizons, the same downward trend is present here, although it does not seem as pronounced. The higher variability can be due to a less important dispersion of forecast performance between variable subsets. The noise in the data likely leads to a reduction in the spread of forecast performance by masking, in part, the added information provided by additional variables. Therefore, subsets of varying cardinalities will have very similar performance for all horizons. In this case, the subset leading to the best results is likely the result of chance rather than the result of distinct properties of the explanatory variables. The higher noise, which entails higher variability, is also possibly responsible for the less evident downward trend in the best subset cardinality for increasing horizons. Because of the noise, or uncertainty, in the variables the best subsets for short horizons present “low” cardinalities when compared to the power measurement case. Since a certain number of variables are nonetheless necessary to attain good forecasts for



(a) Irish case (b) Danish case

FIGURE 5.16: *Improvement in forecasts error of the best subsets per horizon over the subset having the best* NRMSE_{Sub}*. The results are presented for the Irish and Danish case, considering wind speed forecast subsets.*

high horizons, the reduction of cardinality is mechanically constrained by the level of the cardinalities for the lowest horizons.

In these results, it can also be noticed that the cardinalities of the subsets leading to the best per-horizon performance oscillate around the cardinality of the best on-average subset. Since the best on-average subset is different from the best per-horizon subsets, examining the difference in performance between these subsets is necessary. Indeed, if the difference is important, considering the on-average error criterion to determine the best variable subset might not be suitable.

In Figure 5.16 the improvements in terms of NRMSE and NMAE of the best subsets per horizon over the best subset in terms of NRMSE_{Sub} are presented for the Irish and Danish case. The first striking aspect of these results is the amount of improvement observed. In the Irish case the improvement reaches more than 2% for the NRMSE and in the Danish case the maximum improvement reaches 6% of the NRMSE. The improvement values observed for the Irish case are clearly superior to those observed when considering power measurement subsets. With respect to the Danish forecasts, the improvement is of the same overall magnitude as that observed for the power measurement subsets. From a general perspective, the improvement values observed here for both case studies are very similar to those observed for the power measurement subsets for the Danish case. All these curves present the same decreasing improvement with forecast horizon, and barring the high improvements observed in the persistence case for very short-term horizons, the improvement values are comparable.

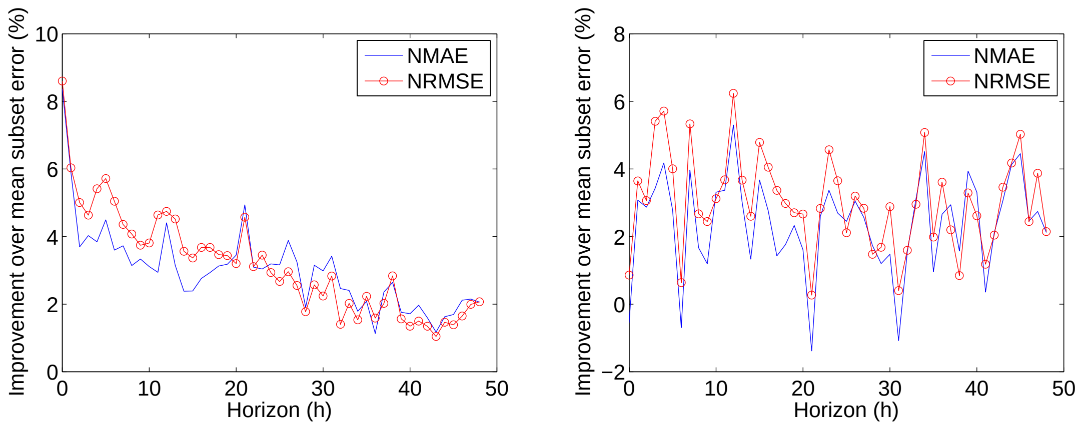
As mentioned in the power measurement analysis, the difference in performance between the best subsets per horizon and the best on-average subset can be attributed to the amount of noise in the explanatory variables. Indeed, in the power measurement case for the Irish data, the information available for horizon 0 is complete and noiseless, and no improvement was observed. In the Danish case, the explanatory variables were noiseless, but the available information was incomplete. Here, the NWP explanatory variables are noisy for both cases. This can explain the fact that, in the Irish case, important improvements are observed although information for all wind farms is available. In the Danish case, the fact that forecasts are not available for all wind farms further degrades the performance of the on-average subset. This can explain the higher improvement values observed for Danish wind speed subsets.

In Figure 5.16, a downward trend of the improvements is apparent, as was the case for the power measurement subsets in the Danish case. This downward trend can possibly be explained by the fact that as the forecast horizon increases the statistical dependence between the wind speed forecast errors increases. Thus, for higher horizons the uncertainty in the explanatory variables not only increases but the redundancy of the variables also increases. Therefore, for higher horizons, different variable combinations lead to forecasts presenting similar error levels. Because of this, the performance difference between the best per-horizon variable subset and the best on-average variable subset is reduced.

Given the noticeable improvement of the best per-horizon variable subsets over the best on-average subset, it can be questioned if this last subset improves over a randomly selected subset. In order to investigate this aspect we examine the improvement, in terms of NRMSE and NMAE, of the best on-average subset over the average NRMSE and NMAE values found for all possible variable combinations. The results are shown in Figure 5.17 for both the Irish and Danish cases. From these results it is clear that opting for the best on-average variable subset yields better results than can be expected by randomly selecting the explanatory variables. For both the Irish and Danish case the improvement is positive for all horizons with regards to the NRMSE criterion. Only in the Danish case do negative improvements in terms of NMAE appear but it should be remembered that the best on-average subset was selected as the best in terms of NMRSE and not in terms of NMAE. Negative improvement values can therefore be considered “normal”, given that the improvement in terms of NRMSE is quite low when the negative NMAE improvements appear.

When comparing the Irish case to the Danish case, the difference in improvement is once again quite clear. In the Irish case, the improvement values range from more than 8% for horizon 0, down to 2% for horizon 48; in the Danish case the improvement values present higher fluctuations and a constant trend around 3%. The less pronounced improvement in the Danish case can be explained by a sharper and narrower distribution of subset errors. Therefore, the best on-average subset only slightly improves over the average subset performance. The reason for this can be that, in the Danish case, forecasts are only provided for a very small portion of wind farms, therefore the “room for improvement” possible by considering different variable subsets is small. There is clearly a limit in the amount of useful information that the explanatory variables can provide. In the Irish case, forecasts are available for all wind farms; therefore it is easier to find a subset that contains relevant variables that are not overly redundant.

The improvement values found here for the NWP based forecast subsets are clearly lower than those found for the persistence forecast subsets examined earlier for horizons



(a) Irish case (b) Danish case

FIGURE 5.17: *Improvement in terms of the NRMSE and the NMAE of the best on-average subset over the mean forecast error computed over all possible subsets. The results are presented for the Irish and Danish case, considering NWP based forecast combinations.*

below 6 hours ahead. For horizons superior to 6 hours ahead the improvements present comparable values. This can possibly be explained by the nature of the errors of the base forecasting models. For horizons below 6 hours ahead — especially for the first three horizons— persistence forecasts present lower error values than the NWP based forecasts; when considered as explanatory variables they are less noisy for horizons below 6 hours ahead and hence choosing an adequate variable subset can lead to significant error reduction. For horizons above 6 hours ahead, persistence forecasts present strongly increasing error levels, while the NWP forecast present more stable, though increasing, error levels. For these horizons, the uncertainty the explanatory variables is much higher which leads to an attenuation of the benefit of choosing the best possible variable subsets. For higher horizons this uncertainty highly penalizes all variable subsets.

From the results presented in this section some conclusions can be drawn with respect to the use of wind speed forecasts as explanatory variables. The most important aspect is that considering too many variables can lead to the degradation of forecast accuracy. The amount of noise, or uncertainty, in the wind speed forecasts hinders the estimation of the forecasting model parameters. This leads to the degradation of the best attainable forecast accuracy when the cardinality of the explanatory variable subset is large. The rule of thumb — considering all available variables — that was formulated in the case of power measurement variables can clearly lead to “sub-optimal” forecasts when considering wind speed forecast. Therefore, implementing an explanatory variable selection method seems necessary, especially when considering the very large number of NWP variables potentially available in a regional forecasting setting.

The second conclusion that can be formulated concerns the error criterion defined in Equation 5.8. This criterion does not lead to the best possible forecasts when considering each forecast horizon independently. As we have shown, the best possible variable subsets noticeably improve forecast performance with respect to the subset chosen as the best subset based on this error criterion. Nevertheless, this criterion does lead to a visible forecast improvement over all horizons when compared to the performance that can be expected from a randomly selected variable subset. Therefore, given the foreseeable complexity of operationally implementing a model whose explanatory variables vary with forecast horizon, the remainder of the work presented here will be carried out considering this criterion as the error criterion for a variable subset.

From a more general perspective, the two conclusions stated above can be extended to all available explanatory variables. Depending on the noise in the variables, their number should be limited in order to ensure that the added noise and the increasing dimension of the forecasting models needed to take them into account do not lead to model parameter estimation problems, or, stated differently, to the degradation of forecast performance. From this, it is clear that in an operational setting where several hundreds of, more or less noisy, variables are available, applying explanatory variable selection methods will be necessary to ensure that the forecasting model presents the best possible forecast performance.

Finally, no single subset leads to the best forecasting performance for all horizons. Therefore, to achieve the best possible forecasts by using models that take forecast horizon into account, variable selection would have to be performed for each horizon, or for sets of successive horizons. This would further increase the time and cost of model tuning as well as the complexity of the forecasting models. Such an approach is of course possible, but in our humble opinion the benefits that could possibly be derived from such an approach are limited given the small forecast improvement that would be achieved. In any case, performing an on-average variable selection is likely to identify a number of subsets that present good performance. From these subsets it might then be possible to derive the subsets that lead to the best forecast performance for each horizon.

5.5 Explanatory variable selection methods

From the results of the study presented in the previous section it is clear that forecast quality strongly depends on the explanatory variables considered by the forecasting model. Selecting the best explanatory variable subset has long been recognized as a major challenge in fields where the number of potential explanatory variables is high and where the relation between variables are nonlinear or imperfectly known. Variable selection has been applied to fields as diverse as spaceship monitoring [126], spam mail detection [127] and hydrology [128]. In this section we examine the possibility of applying variable selection to the regional wind power forecasting problem. The aim is to investigate the possibility of selecting among the available variables the most useful ones, in order to avoid parameter estimation problems and undue model complexity. In the following subsections we present a literature review of variable selection methods that have been proposed in the machine learning literature as well as in other fields. From the methods presented, and considering the behaviour observed in the study presented in the previous section we propose three approaches, the performance of which we examine in the frame of regional wind power forecasting.

5.5.1 Definitions

Variables, Features and Attributes

In the literature the terms variable, feature and attribute are often used when discussing variable selection methods. Often, the terms are used without distinction, but in other cases a clear difference is made. All three terms refer in general to the different “measurable” properties of the process being modelled. In this sense, in a medical setting, where the aim is to determine an illness, the variables, features or attributes that could be used are age, weight and blood pressure.

In particular cases, a distinction is made between variables and features. In such cases variables refer to the input variables of the problem and features refer to “new” variables constructed in some way from the initial input variables [129]. The results presented in Figure 3.16, where the mutual information between the average NWP wind speed and the regional production was computed, can be considered as a case where a feature, the average wind speed, was constructed from available variables. In the work presented hereafter, feature construction or extraction is not considered. To remain consistent with existing literature, and avoid any ambiguity, only the term variable will be used.

The variable selection problem

The variable selection problem can be defined as finding a subset of variables from the original variables of a dataset, such that an induction or prediction algorithm whose learning is performed only on data containing these variables generates a model with the highest possible accuracy on out-of-sample data.

It must be noted that from a theoretical point of view, variable selection is of little interest. Indeed, a Bayes classifier can predict the most probable outcome for a given instance, based on the full probability distribution. In a theoretical setting, where the probability distribution is assumed to be known, the Bayes classifier has the highest possible accuracy [130]. In this setting, the Bayes rule being monotonic, adding explanatory variables cannot decrease its accuracy, therefore restricting the number of explanatory variables appears to be of little use.

However, in a practical setting, the probability distribution is unknown and must be estimated from the available dataset. As discussed in section 4.4, when estimating such a distribution from a finite number of samples, a trade-off exists between the number of parameters to be estimated and the accuracy of the parameter estimation. The number of parameters to be estimated is linked, in part, to the number of explanatory variables considered. By increasing the number of parameters being estimated, the bias of the overall distribution estimation can be reduced. The accuracy of the parameter estimation depends on the size of the training dataset and the amount of noise in the data. By increasing the parameter estimation accuracy, the variance of the overall distribution estimation can be reduced. From this standpoint, reducing the number of input variables can be useful in real-world applications where datasets are finite and often comprise relatively few observations.

Another problem that arises in a practical setting is that of training the prediction or classification model. In most cases, training such a model is a computationally expensive endeavour. Finding the optimal weights of 3-node n-input neural network computing linear threshold functions, has been shown to be an NP-complete problem in some cases [131]; the same has also been shown for other classes of models such as Bayesian networks [132]. The complexity in these cases is shown to be dependent on the n inputs considered. Admitting that $P \neq NP$, the training time for these types of models can be considered to be exponential to the number of inputs n. In this case, reducing the number of inputs considered by a model, can clearly contribute to alleviating the computational burden of model training.

Due to the above problems, the variable selection problem can be more reasonably defined with respect to a particular prediction model. In this way the specificities of the model are taken into account. From this, the problem of variable selection is no longer concerned with finding the subset of variables which lead to the most accurate results, but rather with finding the subset that leads to the most accurate results for a given model. From this, the following definition can be retained [130]:

> Given an inducer $\mathcal{I}$, and a dataset $\mathcal{D}$ with features $X_1, X_2, \ldots, X_n$, from a distribution D over the labelled instance space, an *optimal feature subset*, X_{opt}, is a subset of the features such that the accuracy of the induced classifier $\mathcal{C} = \mathcal{I}(\mathcal{D})$ is maximal.

Note that the above definition derives from the data mining literature. In our case the inducer is the type of model being used to perform regional wind power forecasting. The dataset of features are the available explanatory variables (NWP wind speed, wind direction, measured power, etc), the distribution D over the labelled instance space is the joint distribution of the explanatory variables and the "class labels" are the observed regional wind power productions, the classifier $\mathcal{C}$ is the tuned model. Finally, an optimal variable subset is not necessarily unique, it is possible for different variable subsets to achieve the same accuracy. For instance, if two variables are perfectly correlated, then using one or the other in a subset will not modify the forecast accuracy. This will only lead to two different subsets that have the same accuracy.

Variable Relevance

In the variable subset selection literature the stated aim is often to find a subset of relevant variables. Conceptually, learning can be described as a two step process where one must first decide which features, or variables, should be used to describe a concept and then learning how to combine the information provided by each variable to reach the most accurate decision or estimate. From this point of view, selecting relevant variables and discarding irrelevant variables can be seen as the core aim of variable selection. With this in mind, many different definitions of variable relevance have been proposed in the literature, depending on the specificities of different variable selection algorithms.

Here we present three definitions that apply to our aim. The first two definitions are taken from [130], the third is taken form [133]. Let $S_i = \{X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_m\}$ be the set of all variables except X_i, and s_i a value-assignment to all variables in S_i.

Definition 1 *(Strong relevance). A variable X_i is strongly relevant iff there exists some x_i, y and s_i for which $p(X_i = x_i, S_i = s_i) > 0$ such that*

$$p(Y = y \mid X_i = x_i, S_i = s_i) \neq p(Y = y \mid S_i = s_i)$$

Definition 2 *(Weak relevance). A variable X_i is weakly relevant iff it is not strongly relevant, and there exists a subset of variables S'_i of S_i for which there exists some x_i, y and s'_i for which $p(X_i = x_i, S'_i = s'_i) > 0$ such that*

$$p(Y = y \mid X_i = x_i, S'_i = s_i) \neq p(Y = y \mid S'_i = s'_i)$$

A variable is *relevant* if it is either weakly or strongly relevant; otherwise the variable is irrelevant.

The above definitions are useful for a learning algorithm trying to find which variables are relevant and which are irrelevant. Strongly relevant variables should most often be considered if for no other reason than removing them will introduce some ambiguity into the process being modelled. Weakly relevant variables may or may not be considered. This depends on which other variables are considered. However, as shown in [129, 130], relevance does not imply that the variables will be included in the optimal variable subsets, and likewise, the optimal subset will not include only relevant (strong or weak) variables. This can be due to the "sub-optimality" of the estimation of the underlying distribution or to restrictive model hypotheses.

The above definitions are not specific to a given learning algorithm, and do not imply optimality. In order to better take into account the limitations induced by the practical application of a learning algorithm, the variables' usefulness can be considered as being more significant. In this way, the concept of relevance is to a certain extent replaced with that of variable usefulness:

Definition 3 *(Incremental usefulness). Given a learning algorithm L, a variable subset S, variable X_i is incrementally useful to L with respect to S if the accuracy that L produces using variable set $\{X_i\} \cup S$ is better than the accuracy achieved using just the variable subset S.*

This last definition is particularly useful for variable selection algorithms that explore the variable subset space by incrementally adding or removing variables from their currently examined set. For a more thorough discussion on relevance the reader is referred to [129, 130, 133].

5.5.2 Overview of Variable Selection Methods

The literature on variable selection methods is vast and derived from fields as diverse as statistics, pattern recognition and machine learning. Completely reviewing all available literature would be a daunting task and is outside the scope of the work presented in this thesis. The aim of this section is to provide the important aspects of variable selection algorithms, present some common heuristic search techniques and review some methods that have been proposed in the literature.

Common characteristics of variable selection methods

Most variable selection algorithms are heuristic by nature. Indeed, as shown in subsection 5.3.1, when considering n explanatory variables, there are $2^n - 1$ possible explanatory variable subsets[3] that can be considered to build a predictor. A "brute force" approach, where all possible subsets are evaluated, even with a computationally efficient evaluation function, rapidly becomes unfeasible, even for a small $(20 < n < 100)$ number of explanatory variables. Therefore, heuristic methods are the preferred approach to variable selection.

Given their heuristic nature, most variable selection methods must address a number of aspects in their search through the solution space. The first of these aspects is that of the *search starting point*. A very frequent option is to start with no variables and successively add new variables to the solution being examined. This approach is commonly termed as forward search. The inverse option is also possible, that is: start with all available variables and successively remove variables from the examined solution. This approach is said to proceed backward through the solution space. Finally, another method is to start somewhere in the "middle" of the solution space and explore solutions outward from that point.

The second aspect that the algorithm must address is the *search strategy*, or how to explore the solution space in order to find the best, or approximately best, solution as quickly as possible. As mentioned above, searching through all the solution space can be a very lengthy process, therefore most search strategies aim at examining a limited number of solutions.

The third aspect that must be defined is the solution *evaluation function*. Indeed, the different solutions in the solution space (the different variable subsets) must be evaluated in order to establish which variable subset is the best. There are basically two broad classes of solution evaluation methods: the *filter* methods and the *wrapper* methods [130]. *Filter* methods evaluate the variable subsets based on general characteristics of the data. The relevance and usefulness of the variables is determined by evaluating some property of the variable subset that is supposed to lead to good forecasting performance. In this sense, filter methods are independent of the model to be used as the predictor, the evaluation is based on "ideal" qualities of the subsets. Conversely, *wrapper* methods use the performance obtained by the predictor model, using the variable subset under consideration, as the evaluation criteria. These methods are often more computationally expensive since a predictor must be trained and evaluated for each examined subset. The rationale behind

[3] In much of the variable selection literature the number of possible subsets is reported as being 2^n. Although this can be considered correct since all possible subsets that can be derived from n variables is indeed equal to 2^n, this includes the empty set. In a practical setting, building a predictor with no explanatory variables can hardly be expected to lead to good accuracy, unless the dependent variable is perfectly random, in which case building a predictor makes little sense. Therefore, we consider $2^n - 1$ as being the number of subsets that must actually be examined.

this approach is that filter methods do not specifically take into account the particular biases and hypotheses of a given predictor, hence the variables chosen by a filter might not be the most appropriate for a given type of predictor.

Finally, the fourth issue that the variable selection algorithm must address is the *stopping criterion*. Indeed, as with most heuristic approaches, deciding when to stop the solution evaluation process is necessary. Naturally, the stopping criterion must be adapted to the search strategy employed. Stopping criteria in a forward search can be to stop adding variables if no improvement is observed after the inclusion of a variable. The opposite action can be a criterion for stopping a backward search algorithm.

In the remainder of this section we will mainly concentrate on presenting algorithms that employ different search strategies, and then we will discuss in more detail the filter approaches. We will not particularly discuss wrapper methods as none found in the literature specifically applies to the predictors used in the present work. However, if the reader is interested in existing wrapper methods he can consult [129, 130, 133, 134] and references therein.

Strategies for solution-space exploration

The most common search strategy used in variable selection algorithms is the greedy hill-climbing approach where the algorithm examines the starting point solution and then evaluates possible local changes such as adding or removing a variable. Once the best local change has been found, the algorithm moves on without backtracking if a solution examined further on leads to worse performance. When worse performance is encountered the algorithms usually stop. As mentioned above, either forward or backward selection is used in this type of search strategy. Forward based selection methods can be found in [127, 128, 134–137]. Examples of backward selection algorithms can be found in [138, 139]. The drawback of greedy hill climbing methods is that once a local solution is selected, the algorithm moves on and either adds or removes variables without exploring alternative choices. This can lead to the search becoming "trapped" in a local minimum and missing the optimum variable subset.

To overcome the limitation of greedy algorithms, different heuristic search methods derived from the maximization and artificial intelligence literature are also employed in the variable selection literature. One of these methods is the best first search heuristics. In this type of search, an initial phase similar to that of greedy searching is employed. As the heuristic's name implies, the most promising search path is explored first. When the search path starts leading to worse solutions than the best one found up to that point, the search algorithm is given the possibility of backtracking to a previous solution and exploring a different path from that solution forward. In this way, the algorithm is given greater flexibility and can avoid falling into local minima. If the best first search algorithm is allowed to search for a sufficiently long time it will explore the entire solution space. This heuristic is used in particular in [134].

Another family of heuristic search algorithms that have met with some popularity are genetic algorithms. Genetic algorithms mimic the principles of natural selection. In such algorithms a, "population" of solutions evolves according to the principle of survival of the fittest. The fittest being in the present case the solutions that lead to the best accuracy. In this way, after a certain number of generations, the population is expected to converge toward an optimal solution. The solution's "genome" is usually coded as a fixed length binary string where the value (0 or 1) of each element of the string indicates if a given variable is part of the solution subset or not. The process of evolution is governed by two main operators: *crossover* and *mutation*. Crossover combines parts of the "genomes" of different solutions to create new solutions. This, in a way, mimics the process of reproduction where parent's genes, in our case the variables belonging to a surviving solution, are transmitted to the children solutions. The principle of mutation randomly changes one or several of the "genes" of a solution, that is, the values in the binary string. By applying these two types of operators new solutions can be generated in order to avoid local minima. This approach can be found for example in [126].

Solution evaluation in filter methods

Filter-based variable selection algorithms aim at selecting variable subsets by evaluating general characteristics of the data. Different characteristics can be evaluated. In this section we present some of the characteristics that can be commonly found in the variable selection literature.

One approach is to base the selection on the consistency of the different variable subsets. In such methods a consistency measure is defined and for each variable subset the degree to which the subset violates the consistency measure is translated to a ranking of the variable subsets. The most consistent subset is then considered to be the best subset. Such an approach is described in [139], where the consistency of the variable subsets is determined from the basic property of a function:

> *Property:* If f is a function such that $y_i = f(\mathrm{x}_i)$, $y_j = f(\mathrm{x}_j)$, the $\mathrm{x}_i = \mathrm{x}_j \Rightarrow y_i = y_j, \forall i, j$.

where x_i and x_j are input vectors. A variable subset is consistent if for all observations in the dataset the property is respected. The filter algorithm uses a backward greedy exploration approach. It first considers all available variables and computes the consistency of the available dataset, that is, it counts how many observations χ in the dataset violate the above property. The algorithm then proceeds by computing the consistency values obtained by removing each variable in the subset. If the removal of one variable does not modify the consistency, then that variable is considered redundant or irrelevant and can be removed from the variable subset. This process is repeated until removing any of the remaining variables leads to an increase in the inconsistency of the variable subset. In order to handle noise in the data as well as real valued variables the algorithm uses fuzzy clustering to determine the similarity of the explanatory variable observations and the associated dependent variable observations.

Another family of filter methods uses one learning model to determine the best variable subset for another model. In this sense, this approach is similar to the wrapper approach, the difference being that the filter does not use the intended model but another whose training is selected so as to be less computationally expensive. This approach is used in [137], where a polynomial regression model is used to determine the best variable subset for a multi-layer Perceptron network. In this greedy forward approach, the first variable to be selected in the variable subset is the one that leads to the best accuracy using a given loss function. The variable subset is then expanded by adding to the subset the variable whose addition leads to the highest reduction of the error. The selection process stops if at some point the inclusion of any remaining unselected variable leads to an increase of the polynomial model's error.

One filter family that has met with large success in the variable selection literature is that of information theoretic filters [127, 128, 135, 138, 140, 141]. These methods are based on the use of Shannon's entropy, mutual information and measures derived thereof, to estimate the potential of variable subsets. As presented in subsection 3.4.3, the mutual information $I(Y, X)$ between two variables is a measure of the reduction of the uncertainty concerning one of the variables derived from knowing the other. In the case of variable selection, let us assume that Y is the dependent variable and $\mathrm{X} = \{X_1, \ldots, X_l\}$ is an explanatory variable subset. Then, computing the mutual information $I(Y, \mathrm{X})$ between the dependent variable and the variables in the subset allows to quantify how much knowing the explanatory variables reduces the uncertainty concerning the unknown variable Y. Therefore, the mutual information between the dependent variable and different explanatory variable subsets can be computed and the subset leading to the highest mutual information can be retained as the best subset. In theory this holds, however in a practical setting some remarks can be formulated. The first one being that mutual information does not decrease with the cardinality of the subsets. If a totally irrelevant variable X_{ir} is added to a subset X_a, the mutual information between the new subset $\mathrm{X}_b = \mathrm{X}_a \cup X_{ir}$ and the dependent variable will equal the mutual information found between subset X_a and the dependent variable. Therefore mutual information does not penalize redundant or irrelevant variables. Because of this, most algorithms employ a stopping criterion so that no variables are added into the final set if the difference in mutual information is below a certain threshold. Another inconvenient aspect of mutual information is the difficulty of accurately estimating multi-dimensional probability density functions of continuous variables. The density estimation problem of continuous variables is very similar, if not identical, to that of tuning a learning algorithm. Accurately estimating probability density functions in high dimensional spaces requires a number of observations that is exponentially proportional to the number of dimensions. Therefore, a straightforward application of the mutual information criterion is not applicable if too many continuous variables are available to choose from. This limitation is often overcome by not computing multi-dimensional mutual information, but by using only two-dimensional mutual information measures in the filtering process. An example of this approach can be found in [135].

Clustering filter methods

In the previous paragraphs we have presented variable selection filter methods that build variable subsets by examining some property of the variable subsets or by building variable subsets sequentially based on some statistical measure. Another type of approach reported in the literature is based on variable clustering. The idea behind such methods is to partition all available variables into a number of clusters and then choose one variable from each cluster to form the final variable subset. In this way, the number of redundant variables is reduced. The clustering of the variables is performed using some similarity measure. Variables that are similar are clustered together and variables from one cluster are expected to be very dissimilar from those of another cluster. Several algorithms as well as similarity measures have been proposed in the literature. Similarity measures range from distances such as the Mahalanobis distance to the linear correlation coefficient, to least square projection fit, to combinations of such factors [142]. The clustering algorithms also vary but algorithms where the number of final clusters can be defined are favoured. Often, k-means or other k-nearest neighbour based algorithms are used [138, 142].

The main advantage advocated by the proponents of such methods is the very advantageous computational complexity of clustering methods when compared to other filter methods. As reported in [138], most filter methods have a computational complexity of $\mathcal{O}(l^2)$, where l is the length of the data set, whereas efficient clustering algorithms have a complexity of $\mathcal{O}(D^2)$, where D is the number of available variables. Further, according to this source, since computing linear similarity measures has a complexity of $\mathcal{O}(l)$, a clustering based algorithm which computes linear similarity features will have a computational complexity of $\mathcal{O}(D^2 l)$ which, with most datasets, will lead to much shorter computation times.

5.5.3 Proposed variable selection methods for regional wind power forecasting

As discussed in the previous section many variable selection algorithms have been proposed in the literature. However, to our knowledge, no algorithm presents a clear superiority over other algorithms for all types of datasets. Indeed, from the classification of selection algorithms presented in [143], it appears that the modeller must use his knowledge of the dataset properties as well as the knowledge of the characteristics of the learning model he considers in order to choose *a priori* the selection method to use. These conclusions are further reinforced by the conclusions drawn from the empirical benchmarking of different selection methods presented in [144] for discreet variable selection. In this benchmarking, the authors conclude on the advantage of wrapper methods in terms of accuracy and discuss the advantages of different methods depending on the interdependence existing between explanatory variables.

Clearly selecting an appropriate subset is not trivial, especially when high interdependence exists between variables as is the case in regional forecasting where most of the forecast variables are highly correlated. However, from the literature it appears that wrapper methods are the ones that promise the highest degree of accuracy. However, this type approach is more computationally expensive than filter methods and is only valid for the forecasting model under consideration. Therefore, filters can also be considered as they provide faster computation and are more generic in nature.

In this thesis we have chosen to asses three alternative selection approaches. The first is a filter proposed by Battiti [135], which uses the estimation of mutual information between variables to build variable subsets in a greedy forward selection manner. This method will serve as the reference for the two other methods that we propose. The two other methods are inspired by methods proposed in the literature and try to take advantage of some of the variable subset properties that were put in evidence in the study presented in section 5.4. The first method is a filter method based on variable clustering that is intended to take advantage of the geographical component of the explanatory variables. The third method that we propose is a wrapper approach whose search through the solution space is intended to take advantage of the stability that was observed in the transitions between the best subset of one cardinality to the best subset of higher cardinality.

By examining the performance of these three very different methods, knowledge of the most crucial characteristics of the regional forecasting variable selection problem can possibly be acquired and future paths for improvement can be identified.

Battiti's MIFS algorithm

In [135] Battiti proposed the MIFS (mutual information based feature selection) algorithm. As its name implies, the algorithm is based on the use of mutual information to evaluate the quality of a variable subset. As mentioned in the previous section the idea behind the use of mutual information is to measure the amount of uncertainty reduction concerning the outcome of a variable Y provided by the knowledge of the outcomes of a set of variables X. The idea is then to evaluate the mutual information between explanatory variable subsets and the dependent variable in order to find the subset that provides the highest information on the dependent variable. However, mutual information does not decrease if an irrelevant or redundant variable is added to a variable subset. This can be pernicious to the performance of forecasting models. The most used method is to start with low cardinality subsets and progressively add variables until adding variables does not significantly increase the mutual information between the explanatory variable subset and the dependent variable. Hence, a greedy forward selection algorithm can be used to search for a good variable subset using mutual information as a measure of the variable subset's quality.

However, as mentioned before, computing multi-dimensional mutual information requires estimating multi-dimensional probability density functions. Accurately estimating such densities for dimensions higher than 3 or 4 requires an exponentially increasing number of observations that are seldom found in "real world" applications. To overcome this problem Battiti proposed an algorithm which only requires the computation of mutual information between two variables. This algorithm, presented below - Algorithm 5.1, is particular in that it returns a variable subset of cardinality k defined by the user. In this way, if the user sets k equal to the number of available variables, the algorithm will provide a subset containing all available variables. However, if the order in which the variables are added to the selected variable subset is retained and k is set equal to the number of available variables, then the algorithm provides a variable ranking. This ranking can then be used to determine the subsets of different cardinalities that the algorithm would produce if run for successive values of k. In this way it is possible to produce a number, equal to the number of available variables, of variable subsets of increasing cardinality. It is then a simple matter to test each variable subset and retain the one that leads to the best accuracy.

Algorithm 5.1 Battiti's MIFS algorithm

Set k the number of variables in the selected variable subset
Define set $F = \{X_1, \ldots, X_n\}$ of all explanatory variables X_i.
Define set $S = \emptyset$ of all selected variables.
Find variable $X^*_{i'}$ such that $I(X^*_{i'}, Y) = \max_{X_i \in F} I(X_i, Y)$ where $I(X_i, Y)$ denotes the mutual information between the explanatory variable X_i and the dependent variable Y.
Set $F \leftarrow F \backslash \{X^*_{i'}\}$
Set $S \leftarrow \{X^*_{i'}\}$
repeat
  Find $X_i \in F$ that maximizes $I(X_i, Y) - \beta \sum_{X_j \in S} I(X_i, X_j)$
  Set $F \leftarrow F \backslash \{X_i\}$
  Set $S \leftarrow S \cup \{X_i\}$
until $|S| = k$
Output S

As can be seen in the algorithm, the computation of multi-dimensional mutual information is replaced by the computation of mutual information between individual components. First the variable providing the most information on the dependent variable is selected. The variables that maximize $I(X_i, Y) - \beta \sum_{X_j \in S} I(X_i, X_j)$ are added successively to the selected variable subset. The idea behind this criterion is that a good subset will include variables that are highly informative of the dependent variable Y, while at the same time possessing little mutual information with the variables already included in the selected subset. The information on the dependent variable is maximized by the first term of the criterion $I(X_i, Y)$, and the redundancy of the selected variables is penalized by the second term $-\beta \sum_{X_j \in S} I(X_i, X_j)$. In this way, the algorithm strives to build highly informative subsets while at the same time limiting the amount of redundant variables.

From the above, it is clear that the value of the β parameter will condition the sensitivity of the algorithm to redundant variables. In [135], the author proposes to use values of β between 0.5 and 1. The author observed that for $\beta = 0$ the selected features are, not surprisingly, those that have the highest mutual information with the dependent variable. However, as the value of β increases the algorithm tends to select variables that present lower mutual information with the target variable. From this, it is clear that the algorithm can discriminate between redundant variables. Given the high levels of correlation between explanatory variables in the case of regional wind power forecasting, setting the value of β to relatively high values seems necessary in order to strongly penalize redundancy in the input variables.

To compute the mutual information between variables, we propose to use kernel density estimation in order to estimate the probability density functions of the variables. This is the same method employed in subsection 3.4.3 to compute the mutual information between the wind speed forecasts and regional production. Further details on this method can be found in [145], and further details on kernel density estimation can be found in [100].

The clustering-based filter method

As mentioned earlier, clustering algorithms have been used in the frame of variable selection. Here, we choose to use clustering in order to take advantage of the geographical location from which the variables in the best subsets originate. Indeed, as shown in the study presented in the first part of this chapter, the location from which the variables are selected is not random. The variables found in the best subsets clearly seem to be representative of certain sub-areas in the regions of interest. This representativeness can be explained by the fact that the wind farm's production is governed by the weather situation reigning over the considered region and by the fact that meteorological conditions can be considered to be relatively homogeneous over small areas. For example, the wind speed at one wind farm, or measuring station can be used to determine the wind speed at another site with good accuracy if the two sites are not too distant from one another [44]. Hence, knowing one wind speed in a defined area can be sufficient to determine the wind speed for neighbouring sites.

The idea behind the approach we propose is then to classify the available variables into a number of clusters, based on some similarity measure, and to include the variables closest to the cluster centroids in the selected variable subset, see Algorithm 5.2. This requires defining first a clustering algorithm and deciding which variable similarity measure is most appropriate. In the present case we choose the well-known *k-means* algorithm as the clustering algorithm [146]. As for clustering metrics, several can be used. In the present work we shall restrain our study to the Euclidean distance between the wind farms (the distance computed from the geographical coordinates of the wind farms) and the correlation coefficient between the explanatory variables used by the regional forecasting model.

Algorithm 5.2 Clustering-based filter algorithm

Set k the number of variables in the selected variable subset
Define set $F = \{X_1, \ldots, X_n\}$ of all explanatory variables X_i.
Define set $S = \emptyset$ of all selected variables.
Cluster the variables in F into k clusters C_j using the k-means algorithm
for $i = j$ to k **do**
  Determine X_j the variable closest to the centroid of cluster C_j
  Set $S \leftarrow S \cup \{X_j\}$
end for
Output S

The clustering method will be used in much the same way as the MIFS algorithm. The k-means clustering algorithm is used to classify the wind farms or variables into k clusters. Within each cluster, the farm or variable closest to the cluster centroid is included in the final subset. The clustering algorithm is run n times (where n is the total number of available reference farms or variables) to determine n reference farm or variable subsets of cardinalities 1 to n. When considering reference wind farms, the forecasting model is run n times using the variables provided by the farms in the selected subsets, in this way the best subset can be selected. When considering variable subsets, the forecasting model is also run for the n subsets to determine the one that leads to the best accuracy.

The wrapper method

As discussed above wrapper methods use the forecasting or learning algorithm that is to be used operationally in order to evaluate different explanatory variable subsets. As can be expected, this approach is more computationally expensive than filter methods because the forecasting model must be trained for each variable subset that is tested. The advantage of wrapper approaches however is that the properties of the forecasting model as well as its behaviour toward the intended explanatory variables is directly evaluated, which often leads to better variable subsets being selected. One of the main problems that must be addressed when designing such an approach is the solution space search strategy. Indeed, the expected accuracy depends on the likelihood of the model finding a good or even optimal solution, which is directly related to the search strategy. A strategy that investigates many solutions is likely to find a very good solution but at the cost of a lengthy computation time. Hence, the problem is to determine a strategy that is likely to lead to a good solution while examining the fewest possible solutions.

In the problem of regional wind power forecasting we have seen that the transitions between the best subsets of successive cardinalities are relatively stable when considering one type of explanatory variable. Here we remind the reader that a transition is termed stable if the best subset of cardinality $n+1$ can be derived from the best subset of cardinality n by adding one variable to the subset of cardinality n. Hence if all transition were stable, finding the best subset of cardinality 1 would lead to the best overall subset by a standard greedy search algorithm. However, we have seen that transitions, although very stable, are not perfectly so. Therefore, a “traditional” greedy forward approach has a great chance of missing the best subset especially since the highest instability occurs for low dimensionality transitions. To overcome this problem we propose a greedy forward selection approach that examines more solutions but has a higher chance of finding the best solution. The proposed search method is presented in algorithm 5.3.

Algorithm 5.3 Stability-based wrapper algorithm

```
Evaluate the accuracy of all the 1-variable subsets
for i = 3 to n do
    Determine SC_{i-1} the h best subsets of cardinality i − 1.
    Evaluate the accuracy attained when using the variable subsets of cardinality i derived
    from the subsets in SC_{i-1}.
end for
Select the best combination from all those computed.
```

The idea behind the search procedure is to determine the set SC_i of m best subsets of cardinality i by applying some error measure to the results obtained by the forecasting model when using the evaluated variable subsets as inputs. Once the set SC_i has been determined, the subsets of cardinality $i + 1$ whose transition is perfectly stable with the subsets in SC_i are evaluated. In this way only the subsets deriving from the h best subsets of lower cardinality are examined. This leads to a greater number of subsets being examined than with more traditional greedy forward selection algorithms. The actual number of subsets that are examined depends on the value of h and the structure of the solution space. The reader must have noticed that the algorithm does not use a stopping criterion. We have omitted the use of one in order to more fully explore the solution space. However, in a setting where the initial dataset contains many variables, a stopping criterion such as no improvement after a certain number of cardinality increases can be used to reduce the number of examined solutions.

In our work the error criterion is the average NRMSE error over all forecast horizons defined in Equation 5.8. Another important parameter that must be set in this algorithm is the h parameter. This parameter governs to a certain extent the number of solutions that will be examined by the algorithm. In the present case, the transitions are rather stable so a small value of h is likely to lead to good results. However if the transitions are more unstable, the h parameter can be increased so that a greater number of solutions is examined and the likelihood of finding a good or optimal solution is increased. Furthermore, different values of h could be used for the different subset cardinalities. Indeed, there are very few 2 and $n-1$ variable subsets whereas the number of possible subsets for cardinalities near $\frac{n}{2}$ is very high. Therefore, the h parameter could be set to vary proportionally to the number C_i^n of possible subsets of cardinality i. This however would augment the number of explored solutions. In the present case, the instability is high for low cardinalities, and low for higher cardinality transitions. Hence examining the subsets derived from a fixed value of h will lead to the examination of a high percentage of subsets of low cardinalities and a decreasing proportion of subsets of higher cardinality, this will allow the algorithm to thoroughly investigate the higher instability transitions of low cardinalities and limit the search once more stable transitions are encountered.

5.6 Evaluation of the proposed variable selection methods

In this section we present the results of the three variable selection methods presented in the previous section. We first present the evaluation framework applied to each selection method and we then analyze the results.

5.6.1 Scope of the study

In order to analyze the variable selection methods we must first define an evaluation framework. Evaluating the performance of variable selection methods requires the same approach used when examining the performance of forecasting models: cross-validation. The initial dataset must be divided into two separate datasets. The selection algorithms are first run on one dataset to determine the best subsets then the results are evaluated on a testing set to see if the methods "generalize" well. This approach however is not necessary for all the methods proposed in the previous section.

Cross-validation will be used to evaluate the MIFS and wrapper method as well as the clustering method when using correlation between variables as the distance measure. This however is unnecessary for the clustering approach when using the distance between wind farms. When using the distance between wind farms the clustering method does not use the values contained in the dataset to determine the best subset, but rather a characteristic of the variables that is to some extent independent of the information they contain. Therefore, for this method only the results obtained on the testing set will be examined.

In order to allow comparisons to be made between the different approaches, the same learning and testing sets will be used for all methods. These will be the same as those used in the variable subset selection study presented in section 5.4. In this way the results of the methods will also be comparable to those found in the exhaustive search that was carried out therein. In the case of the MIFS algorithm, the mutual information will be computed only on the learning set. For the wrapper method, which requires the use of the forecasting model, the main learning set will be further divided into a learning and testing set so that appropriate cross validation can be applied. These sets represent respectively 20% and 80% of the learning set.

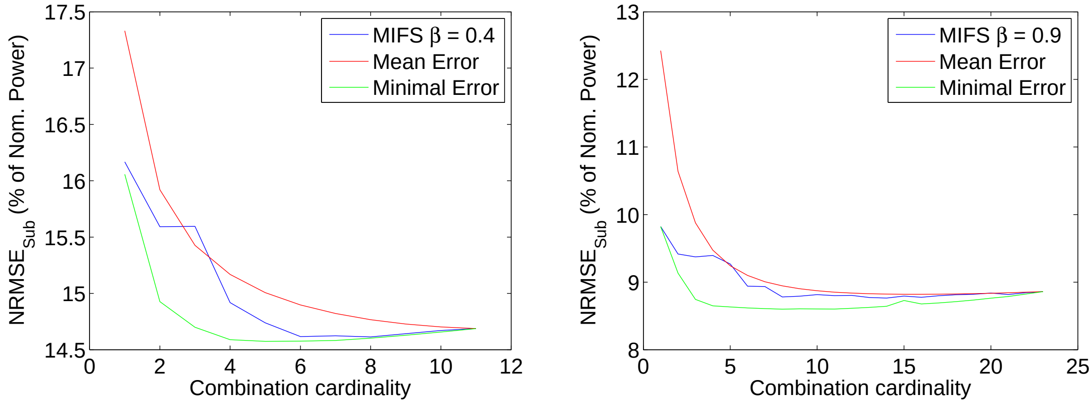
With respect to the computation of the different criteria used by the methods, all forecast horizons will be taken into account. For correlation as well as mutual information, all horizons will be considered as being part of the same variable. In this way these criteria will be consistent with the average NRMSE criterion used in the wrapper method. By preserving this consistency, the results of the methods will also be comparable to the exhaustive search conducted in the study presented in section 5.4.

Finally, the methods will be tested for the case of wind speed forecast selection. This is the case where variable selection is most necessary since for both case studies (Ireland and Denmark) "sub-optimal" subsets appeared once a certain subset cardinality is exceeded.

5.6.2 Results for the MIFS Algorithm

In order to examine the performance of the MIFS algorithm two aspects are considered. The first is the "raw" performance of the algorithm for a given case study. The second aspect is the influence of the β parameter on the performance of the selection algorithm.

To evaluate the performance of the selection algorithm we compare the performance of the selected variable subsets to the best performance obtained for each subset cardinality. Further, since the selection algorithms examined here are heuristic, we also compare the performance of the selected variable subsets to the average performance of the variable subsets of a given cardinality. The results for the Irish and Danish case studies are presented in Figure 5.18. These are the results that lead to the best performance for given values of β. In other words they are the best results that the method could obtain if the correct value of β to be used were known *a priori.*



(a) Irish case

(b) Danish case

FIGURE 5.18: *Performance of the MIFS algorithm for the case of Ireland and Jutland. The* NRMSE_{Sub} *is plotted against the variable the subset cardinality. In the Irish case the* β *parameter is set to 0.4 and in the Danish case the* β *parameter is set to 0.9.*

The first striking aspect of the results is that the MIFS algorithm does not lead to optimum variable subsets for either case study, with exception of the one-variable subset in the Danish case. However, for most cardinalities, the subsets selected by the algorithm lead to better performance than that which could be expected by randomly choosing the explanatory variables. The only cardinality for which the algorithm leads to worse performance is cardinality 3 in the Irish case and cardinality 5 in the Danish case. Clearly using this selection algorithm is useful and can lead to the selection of variable subsets that are better than randomly selected ones.

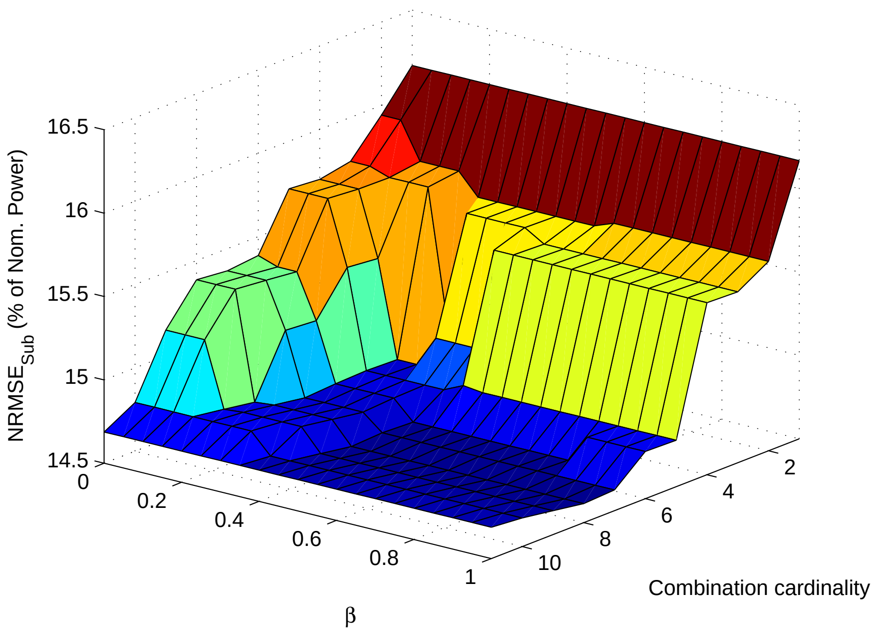
The second aspect that can be noticed from these results is the very similar shape of the performance curves obtained for both cases. The error decreases for the first few subset cardinalities and then reaches a plateau where it remains constant or even increases for some cardinalities before decreasing again. Further, the second error decrease takes place when the subset cardinality nears the cardinality having led to the best overall performance: 5 for the Irish case and 8 for the Danish case. Indeed, for the Danish case, the best subset selected by the algorithm has the same cardinality as the bet possible subset.

The peculiar shape of the error curves can possibly be linked to the constant nature of the β parameter. In the selection algorithm the value assigned to β does not vary with the cardinality of the subsets being considered. This can lead to the inclusion of variables that are not redundant after only a few cardinalities. The added variables are not redundant but do not provided the forecasting model with much useful information. After a certain number of variables have been included, the algorithm has no other choice but to include what it sees as redundant variables, which nevertheless provide useful information to the forecasting model. A possible improvement on this algorithm could then be to use a varying value of β in order to modify the sensitivity of the evaluation function to redundant data. In this way, very relevant variables might be included with a certain tolerance to redundancy up to a certain point and then redundancy could be more strongly penalized.

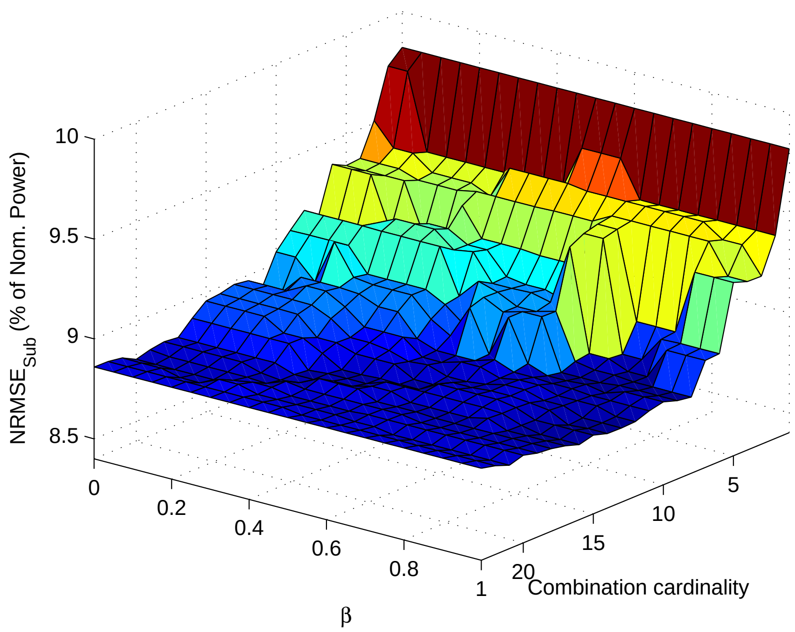
To further examine the impact of the β parameter we compare the performance obtained for a wide range of values of β. The results of this comparison are shown in Figure 5.19, where the performance of the selected variable subsets is plotted against their cardinality and the values of β. From these results several conclusions can be drawn with respect to the influence of the β parameter, as well as some general guidelines for selecting a value, *a priori*, depending on the properties of the variables under consideration.

The first conclusion that can be drawn is that the value of the β parameter heavily influences the results of the selection algorithm. From a general standpoint, very low values of β lead to the worst performance for all subset cardinalities. As the value of β increases, the tolerance of the selection method toward redundant variables decreases, and the results tend to improve. However, after a certain threshold, the performance of the model starts decreasing, especially for the medium range cardinalities that are close to the "optimal" cardinality. This clearly illustrates the negative influence that exclusively selecting the most relevant or the least redundant variables, can have on the forecasting model performance. When β is small, the selection algorithm favours the selection of the most relevant variables without regard to their redundancy. When β nears 1, the selection algorithm penalizes redundancy more than it favours relevance so that the primarily selected variables are those that present the least redundancy between themselves, with limited regard to their relevance. As can be expected the "best" results are found for values that provide a certain comprise between the importance of variable relevance and variable redundancy. This is apparent in the Irish case, where the best result was found using a β value of 0.4. However this could seem contradictory with the fact that best results in the Danish case was obtained using $\beta = 0.9$. This leads us to the second conclusion that can be drawn from these results: the relation that seems to exist between variable dependency and the values of β that lead to the best results.

Clearly, the impact of the beta parameter varies between the two case studies. In the Irish case, the results very rapidly improve as β increases. For values of β between 0.3 and 0.6 the performance of the selected subsets is the best and then, for higher values of β, the performance decreases. In the Danish case, the improvement is much less pronounced, with a very gradual improvement of the results with increasing values of β. This difference can be linked to the difference in the relations existing between the variables in both cases. The correlation between the explanatory variables is much higher in the Danish case than in the Irish case. Hence, the redundancy of the variables is much higher, which means that when it is neglected by the selection algorithm (low values of β) very redundant variable subsets will be selected. In the Irish case, where the dependence between explanatory variables is lesser, lower values of β prove discriminating enough to derive good variable subsets. Therefore, defining what value of β to use with the MIFS algorithm can benefit from an initial examination of the relations existing between the explanatory variables. It must however be noted that, in the present case, almost all explanatory variables present very high correlations with the dependent variable, therefore the strength of the relation between the explanatory variables can be used as an indication to the best range of β values to use. If the strength of the relation between the explanatory variables and the dependent variable



(a) Irish case



(b) Danish case

FIGURE 5.19: *Influence of the β parameter on the performance of the MIFS algorithm for the case of Ireland and Jutland. The value of the β parameter ranges from 0 to 1, with 0.05 increments.*

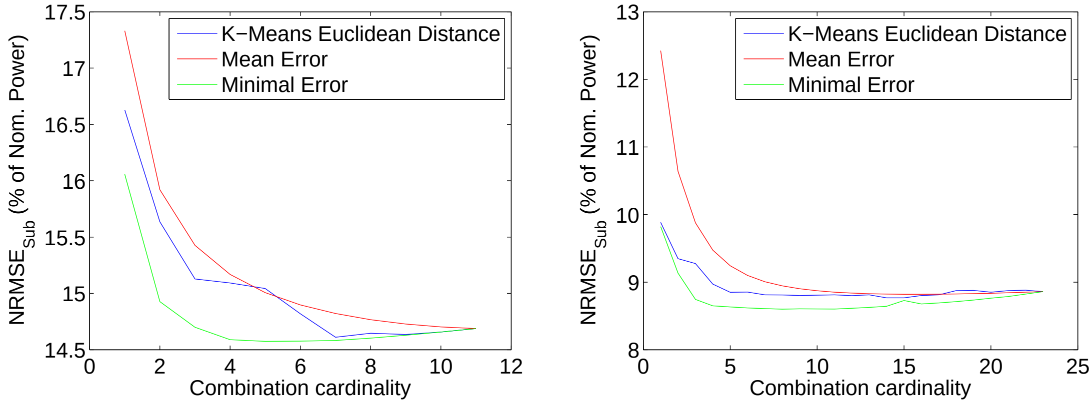
were more heterogeneous, providing an indication of the possibly best range of β values to use could prove to be much more complicated.

5.6.3 Results for the Clustering Method

Before proceeding with the analysis of the clustering approach we remind the reader that this approach is an unsupervised selection method. The quality of the variable subsets is not directly evaluated. Rather, the subsets are chosen so that the selected variables are as dissimilar as possible, the aim here is to select the least redundant variables with no direct estimation of the variables' relevance. In this section two distances were used. In the first approach the geographical coordinates of the reference farms where used to compute the Euclidean distance between wind farms. This distance was then considered as the distance existing between the variables. This approach is expected to capture the geographical representativeness of the reference wind farms. The second distance that is examined is the correlation coefficient between the single-wind-farm forecasts provided for each reference wind farm. This approach is expected to better capture the redundancy of the variables given that it is based on the variables themselves rather than on the geographical location for which they are provided.

The results of the clustering using the Euclidean distance between reference wind farms are presented in Figure 5.20, for both case studies. The average error for the selected variable subsets are compared to the average and minimal error obtained for subsets of the same cardinality. For both case studies the selected subsets present good results with error values below the average error for almost all cardinalities. It must be noted however that for the Irish case the only cardinality for which the clustering algorithm leads to worse than average performance is the cardinality for which the minimal error is achieved. Also, in the Danish case, the performance of the selected subsets is worse than the average for very high cardinalities. This behaviour can be explained by the fact that the physical distance between wind farms is not sufficient to thoroughly discriminate between variables. As seen in the characterization of the regional production in chapter 3, distance can explain part of the relation existing between variables but it does not fully do so. Therefore, the algorithm cannot perfectly select the best subsets.

Although basing the clustering on the Euclidean distance between wind farms does not lead to the best variable subsets, it must be noted that the selected subsets lead to forecast performance that is better than average. When compared to the results obtained using the MIFS algorithm, the subsets obtained with the clustering algorithm compare favourably. This is especially apparent for low cardinalities (below 5 for the Irish case and below 10 for Danish case) where the variable subsets selected by the clustering approach outperform those selected by the MIFS algorithm. This is rather remarkable given that the only information used to select the variables is their geographical coordinates. From a practical angle this is far from negligible since clustering using the Euclidean distance in a two-dimensional space is computationally inexpensive when compared to the MIFS approach, which requires multidimensional density estimation. This method could well be used to determine "reasonable" variable subsets when the number of reference wind farms is high



(a) Irish case (b) Danish case

FIGURE 5.20: *Performance of the clustering based selection, using the Euclidean distance between wind farms, for both case studies. The* NRMSE_{Sub} *is plotted against the variable the subset cardinality.*

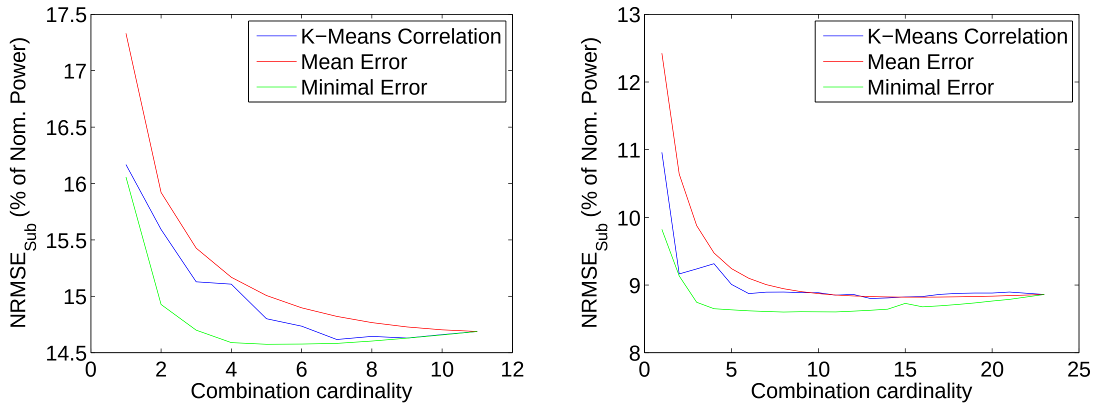
or as a preliminary approach to determine starting positions for more advanced solution space exploration methods.

Clearly, the fact that the relation between variables depends in part on the distance separating their points of origin allows the clustering algorithm using Euclidean distance to select better-than-average subsets in most cases.

In order to further examine the possibilities of this unsupervised approach it is necessary to compare the above results with those found when using a distance that directly measures the dependence between variables. For this purpose we propose to investigate the use of the correlation coefficient as the measure of dependence between explanatory variables. Although this coefficient only measures the linear dependence between variables, we use it because its computation is not too computationally expensive and it provides a good estimate of the dependence between the variables under consideration here.

In Figure 5.21 the results obtained with the clustering approach using the correlation between variables as the distance measure are presented for both case studies. There is an apparent improvement with respect to the results obtained using the Euclidean distance. The selected variable subsets for the case of Ireland lead to better than average forecasting error for all cardinalities. For the Danish case, the results are better than average for almost all horizons with worse than average results only for high subset cardinalities. We can notice that there is also a slight degradation in the error reduction for cardinality 4 for both cases. This same behaviour was observed with the Euclidean distance measure and the MIFS algorithm. The presence of this peak can possibly be due to a region in the solution space where the evaluation of the criteria used here leads to high uncertainty with respect to the performance obtained using the selected variables subsets. This can possibly be due to the combination of the high number of possible subsets and the relatively wide error envelope observed for these low cardinalities.

When comparing the results obtained using correlation to those obtained using the Eu-

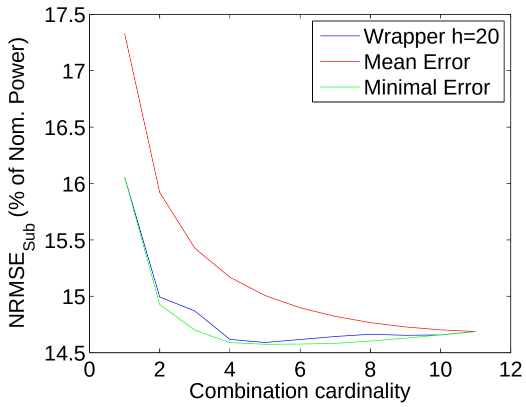


(a) Irish case (b) Danish case

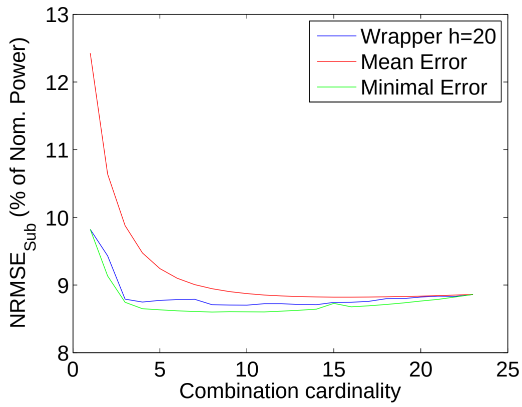
FIGURE 5.21: *Performance of the clustering based selection, using the correlation as the distance between wind farm forecasts, for both case studies. The* NRMSE_{Sub} *is plotted against the variable the subset cardinality.*

clidean distance, the results obtained using the former clearly outperform those obtained using the latter measure. Clearly a more discriminating variable dependence criterion leads to better variable subsets. The variable dependence, which is implicitly present in the distance between the variables' origins, does not provide an accurate enough estimate of this dependence, whereas correlation, although it only considers linear dependence, better captures the strength of the relation between variables. From this it can be supposed that the results obtained using the clustering approach could be further improved by using a distance measure that also considers nonlinear dependencies between variables. The clustering method could be used with a mutual information based distance measure. With such a measure, the definition of cluster centroids is likely to be more complex than the average of the variables used in the clustering approach. Indeed, to fully take advantage of a nonlinear measure, it seems necessary to establish a nonlinear centroid derivation procedure that preserves the nonlinear dependence present between variables.

A further aspect that can be noticed is the difference in performance that is apparent between the Danish and Irish cases for both distance measures. The results are consistently better for the Irish case. This can be linked to the difference in the spread of the dependence values that exists between both case studies. In the Danish case the dependence values lie in a narrower range, than in the Irish case. Clearly the unsupervised approach used here is less effective for cases where the difference between dependence values is small. The clustering algorithm selects those variables that are most representative of existing variable clusters. The algorithm is only effective if the variables can be significantly divided into different clusters. If the variables can only be divided into a small number of significant clusters then, when the algorithm is asked to partition the variables into a higher number of clusters, the partition is no longer based on the significant structure of the variable space. After a certain variable subset cardinality is exceeded, part of the partition is based on differences in distance between variables that are the result of noise in the data rather than the



(a) Irish case



(b) Danish case

FIGURE 5.22: *Performance of the wrapper using the correlation as the distance between wind farm forecasts, for both case studies. The* NRMSE_{Sub} *is plotted against the variable the subset cardinality.*

result of significant variations in the dependencies between variables.

5.6.4 Results for the Wrapper Algorithm

In this section we examine the results obtained using the wrapper approach. We will investigate two aspects of the method. First we will discuss its performance using a single value of the h parameter. We remind the reader that the h parameter defines, for each iteration, the number of variable subsets considered for a given cardinality n to generate the subsets of cardinality $n + 1$ that will be examined by the algorithm. Then, since the results obtained by the algorithm depend in part on the number of examined subsets and since this is linked, also in part, to the value of the h parameter, we will examine the evolution of the results as the value of the h parameter varies. This will also allow us to illustrate the link between the number of examined subsets and the h parameter.

In order to examine the performance of the algorithm we have retained the results obtained for $h = 20$. This value ensures that a sufficiently large number of subsets are examined, so that the algorithm has a good chance of finding good solutions. Figure 5.22 presents the results obtained on the testing set using the selected subsets. These are compared to those obtained using the "optimal" subsets and to the mean error for each subset cardinality.

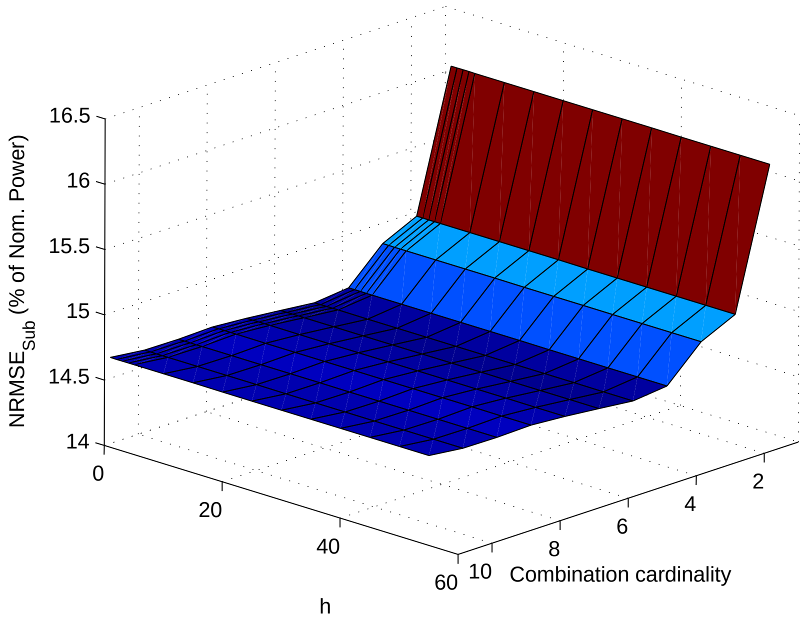
When compared to the results obtained with the previously examined methods it is clear that the results of the wrapper method are quite superior. For both case studies the selected subsets lead to results that are very close to the optimal results, for almost all cardinalities. Moreover, the same behaviour observed with the other methods is apparent with this method: namely the relative performance degradation for cardinalities between 2 and 4. This indicates that for these cardinalities the uncertainty is high with regard to relation between the estimated performance of a subset and the observed performance on the testing set. Although the wrapper method uses the forecasting model as the performance estimator, it fails to accurately predict the behaviour of the subsets on the testing set for small cardinalities. This can be attributed to two factors: the non-stationarity of the process under consideration and the variability in the performance of variable subsets of small cardinality. The non-stationarity of the process renders cross validation difficult to perform. Using one part of the dataset to select the variables and testing the performance of these subsets on a subsequent part of the dataset can lead to a miss-estimation of the performance when the process being model is not stationary. This is one of the most discussed aspects of wrapper method design [130]. Nevertheless, given the difficulty of defining robust cross validation schemes, we do not further examine this aspect, especially given the satisfactory results obtained with this method.

The wrapper method leads to the selection of variable subsets that share the same cardinality as the optimal subsets, although the optimal subsets are not selected. This aspect is interesting since it will lead the modeller to choose a variable subset having a reasonable number of subsets. The selection methods examined earlier often led the modeller to choose variable subsets of cardinalities higher than the optimal cardinality. In this way, the wrapper method can help the modeller to choose a reasonable subset cardinality. This aspect is important given the very negative effects that high input cardinalities can have on out-of-sample model performance. From a more practical point of view, the model resulting from the variable subsets selected by the wrapper method will require fewer variables thus reducing the cost of data acquisition and reducing the complexity of an operational setup.

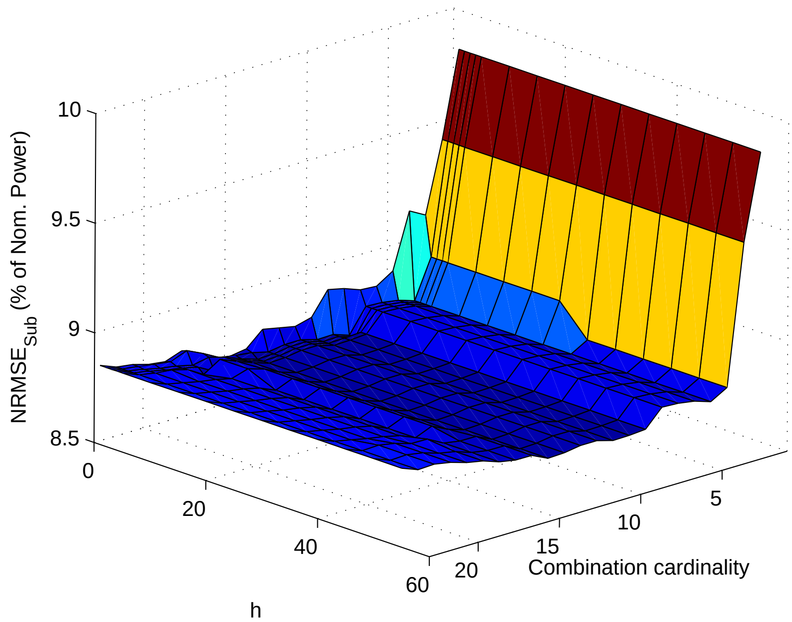
The second important aspect of the wrapper method after its performance, is the value that should be assigned to the h parameter. As mentioned above, the number of subsets that the wrapper will examine depends in part on the value of this parameter. Also, depending on the transition stability between variable subsets of successive cardinality, the results obtained by the wrapper will vary as the value assigned to the h parameter is increased. Hence, two questions that must be addressed when setting the h parameter are : what value of h should be used in order to ensure that reasonably accurate subsets are found and how does the number of investigated subsets increase as the value assigned to h increases.

We present below the results obtained for successive values of h. In Figure 5.23 several runs were performed for values of h between 1 and 55. For values between 1 and 5, all values were tested, then every 5 values were examined up to 55. From the figure it is apparent that even for small values of h, the wrapper approach finds the subsets that it will consider as best. This is particularly clear for the Irish case where the value assigned to h has no impact on the selected subsets. The variable subsets selected by the wrapper are the same for $h = 1$ through $h = 55$. For this case, the stability between variable subset transitions is such that the wrapper does not need to examine very many subsets in order to find the subsets it considers as being the best. For the Danish case the impact of the h parameter is more apparent. For very small values of h (1 to 5) the performance obtained by the wrapper varies importantly. However, for higher values the selected subsets are almost always the same with only minor difference as the value of h increases.

The difference between the behaviours observed for both case studies can be attributed



(a) Irish case



(b) Danish case

FIGURE 5.23: *Evolution of the performance of the wrapper based selection for successive values of h.*

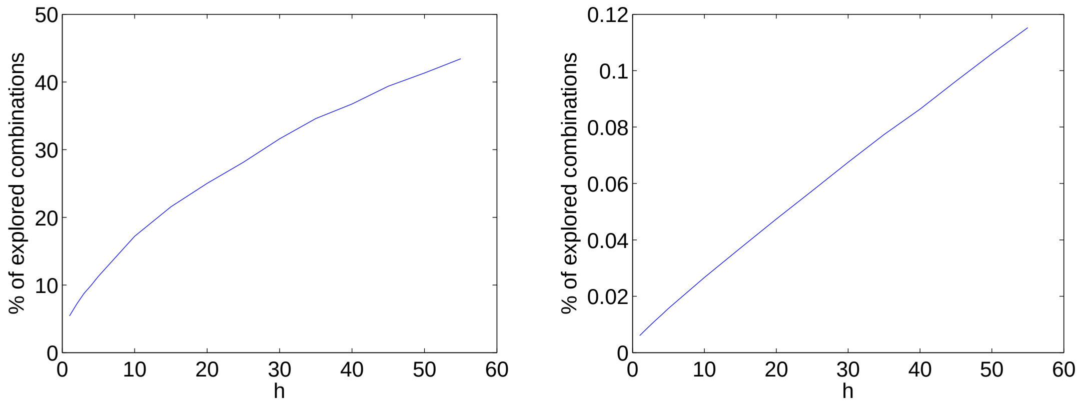
to the number of examined subsets. For small values of h the relative number of subsets examined in the Irish case is high, whereas the relative number of subsets examined for the Danish case is very small for low values of h. Because of this, the wrapper algorithm finds what it deems to be the best subsets for $h = 1$ in the Irish case, whereas a value of h nearer to 10 is necessary before the wrapper finds relatively stable solutions in the Danish case. This phenomenon is further amplified by the transition stability between the subsets that the wrapper considers as being the best. In the Irish case the transitions are stable, whereas in the Danish case some instability is present as can be observed in Figure 5.23(b) where new “best” subsets are found when the wrapper is run with $h = 30$ instead of $h = 25$. From these results it appears that running the wrapper with an h set to values above 10 yields good results.

The relation between the value assigned to h and the number of examined subsets clearly plays a role in the ability of the wrapper method to find good variable subsets. As mentioned earlier the value of h conditions in part the number of subsets that will be examined since it determines the number of subsets that are retained at given iteration to determine those which will be examined during the next iteration. However, no direct relation can be found since the two subsets of a given cardinality that share all but one variable, will both lead to the examination of an identical subset of superior cardinality. Therefore, if the h best subsets of one cardinality are all similar, the number of examined subsets of the next cardinality will be lower than if the h best subsets are very different.

In Figure 5.24 the number of explored subsets is plotted against the value of h for both case studies. Apart from the evident difference in the relative number of explored subsets — in the Danish case this is nearly two orders of magnitude smaller than the number explored in the Irish case — it is apparent that the relation presents some differences between both cases. This difference can be linked to differences in the similarity of the h best subsets found for each cardinality between both case studies. These subsets present higher similarity in the Irish case hence by increasing h only a small number of additional subsets are explored. This does not seem to be the case for the Danish variables where an increase in h leads to an almost proportional increase in the number of explored cases.

From this difference it is apparent that in the Danish case the performance of widely differing subsets is similar. In the Irish case the best performance is obtained by subsets that share an important number of variables. This again can be linked to the nature of the relation between the variables in both cases. In the Irish case, the variables present lower dependencies between themselves and provide different information to the forecasting model. Therefore a given set of variables is necessary in order to obtained good forecasts. If these variables are included in a subset then the performance will be good without regard to the other variables included in the subset. In the Danish case, the dependency between variables is much stronger, hence they all provide very similar information to the forecasting model. For this reason very different variable subsets can present very similar performance.

From these results it is apparent that knowing the strength of the variable dependency can help in setting the value of h. In a case where the dependency between variables is low,



(a) Irish case (b) Danish case

FIGURE 5.24: *Number of subsets examined by the wrapper algorithm as a function of the h parameter. The number of explored subsets is provided as a percentage of all possible subsets for each case study.*

the h parameter can be set to low values (5 to 20) without too much risk of missing the best variable subsets. In cases where variable dependency is higher, the h parameter should be set to higher values (above 20) in order to ensure that a sufficient number of different subsets will be examined and that good subsets are found. Another aspect that should also be taken into account is the size of the solution space. Here the size of the solution space is taken to mean the number of possible variable subsets. The h parameter should also be set so that a sufficient number of subsets are examined from an absolute standpoint.

5.7 Conclusions

In this chapter we have examined the influence that variable selection can have on forecast accuracy in the frame of regional wind power forecasting. We first proposed a framework for the evaluation of the impact of variable subsets on regional forecasting model accuracy. We applied this framework to two types of variables: measured power and NWP wind speeds. From the results of this study it is apparent that different types of variables have a different bearing on the forecast accuracy that can be expected from a model. More specifically, we showed that noisy, or uncertain, variables are the ones that have the greatest impact on forecast accuracy. From these findings we then examined the possibility of using variable selection methods in order to select subsets of explanatory variables that can lead to the best performance without exposing the models to the risk of parameter misestimation.

We identified two main types of variables that have a non-negligible impact on forecast performance, namely measured variables and forecast variables. We examined these two types of variables separately. From the results of the study it appears that measured power carries very little noise and can be used almost without selection. This is especially true in cases where measured power is available for all wind farms. However if power measurements are only provided for a few reference wind farms, then some selection can be necessary in order to avoid model parameter estimation problems. With regard to forecast wind speeds, these are much noisier, and greater care should be taken in their selection. Indeed, since these variables are forecast they present higher correlation amongst themselves and higher error correlations than the measured variables. Because of this, they are more redundant and can degrade model performance more quickly than measured power when too many are considered by a forecasting model.

The main finding from the study of forecast wind speeds is that a relatively few variables are necessary in order to attain good results. Few variables are necessary to accurately characterize the meteorological situation over a region. From this, it appears that the best forecasting performance is obtained when a model considers a limited number of forecast variables. This contradicts to some extent the general idea suggested in the literature [24, 93] that having more variables renders forecasting easier. However this also implies that performing variable selection is necessary in order to avoid high input dimensions which can lead to negative effects on model performance. From a more practical perspective, this also implies that simpler data acquisition solutions are necessary to obtain reliable forecasts.

From these findings the possibility of using variable selection methods to reduce the number of variables used was investigated. The study primarily aimed at selecting from the forecast wind speed variables since from the results referred to above, these variables are those that present the highest likelihood of exposing the model to the curse of dimensionality. Although the performance of the methods could be examined in a more general setting by considering all types of variables, this would require a great amount of computation time and would not necessarily lead to insight on the properties of the selection methods being considered. Such a study could be conducted in future work in order to understand possible interactions between different types of variables.

In this chapter we have examined three variable selection methods; the state-of-the-art MIFS algorithm, and two original methods: a clustering based algorithm and a wrapper algorithm. As we have seen all these methods permit the selection of variable subsets that lead to better than average forecasting performance. Although the methods have differing levels of performance, they present different complexities so that a trade-off is necessary between the performance of the selected variable subsets and the computation time. Further, although these methods present good performance they can be further refined in order to more fully take into account of the specificities both of the regional forecasting problem and the type of forecasting model under consideration.

From the results presented in this chapter it is clear that the clustering based method that uses the distance between wind farms to determine the variable clusters is the method that leads to the poorest performance. However, this method is also the simplest of those examined. Indeed, it requires very simple computation and does not require any measured or forecast data to be used. The method relies on the fact that the dependency between explanatory variables is related to distance separating their point of origin. In this sense this method can be used prior to any data collection. Moreover, the method could be used by not only considering the wind farms or model grid points available but could also include wind farms for which no data is available. In this way the representativeness of the reference wind farms or NWP model grid points with respect to all the wind farms in the region could be taken into account in a more complete way.

The next best method is the clustering algorithm, which uses correlation as the measure of distance between variables. This method is more accurate than the one based on distance between wind farms but it requires more computation time. Nonetheless, the computation of correlation is simple and the method can be easily implemented.

In terms of computation time the next most expensive method is the MIFS algorithm. This algorithm presents good performance and has the advantage of being based on a measure that can be applied to any variable selection problem, namely mutual information. However, the evaluation of mutual information is far from trivial especially in the frame of continuous variables where probability density functions must be estimated. This method provides good results and could be enhanced by taking into account mutual information estimates computed over more than two variables. This however would increase computation time as well as expose the modeller to the difficulty of accurately estimating multi-dimensional probability density functions.

Finally, the method that requires the most computation time is the wrapper method. This method leads to the best results but it is constrained to a given forecasting model. Although this can be considered as a limitation, the heuristic on which the search of the solution space is based can be applied to any model. Further, if different models are implemented so as to be compatible with a core wrapper function, then this approach can be applied to different models with little development overhead.

Apart from the MIFS algorithm, which was adapted verbatim from the literature, the clustering and wrapper methods were developed in this thesis taking into account the specificities of the regional wind power forecasting problem. From the results it is clear that selecting the input variables is possible without examining all the possible subsets. However, there is still room for improvement and other properties of the regional forecasting problem could be taken into account in order to refine the three proposed heuristics. Another aspect that would prove useful would be to extend the study to other case studies so that the impact of the heuristic-specific parameters (β for the MIFS, distance measure for the clustering approach and h for the wrapper method) on the selected variable performance can be better understood and further guidelines provided to model developers.

CHAPTER 6

General Conclusions

6.1 Overall conclusions and contribution

When the work presented in this thesis began some years ago, only a few power systems were confronted with important wind power penetration. In the available literature at the time few regional prediction models were reported, describing specific solutions put into practical use in some power systems. At the time most of the work on regional forecasting concentrated on establishing models and examining the benefits of production smoothing on forecast error reduction. No guidelines were available for other modellers wishing to implement regional forecasting models. In the present thesis regional wind power forecasting has been investigated from a general standpoint. We have shown that regional forecasting cases share common properties and that these properties can be exploited when designing regional forecasting solutions. From the modelling perspective we have shown that little difference exists between different modelling strategies and that different learning algorithms have comparable performance if their modelling capacity is similar. Furthermore, we have shown that the idea of combining very numerous single wind farm forecasts or of exploiting a growing number of available explanatory variables does not lead to an indefinite increase in forecast accuracy; actually only a few, well selected, explanatory variables are necessary in order to adequately forecast regional production.

In order to study regional forecasting from a more general perspective, characterizing the problem was necessary. We examined the statistical characteristics of regional production and showed that regional production is a non-stationary process. The non-stationary nature of regional production can be dealt with by using adaptive forecasting models that periodically re-estimate their parameters. We also showed that regional production presents high autocorrelation. For such process and for short-term horizons, autoregressive models, including persistence, will therefore lead to acceptable forecasting accuracy. This high autocorrelation stems in part from the statistical smoothing of aggregated wind farm productions. To characterize the smoothing of regional production we studied, from a theoretical and experimental angle, the factors that influence it. From this, we concluded that statistical smoothing is proportional to region size and the dispersion of the wind farm within the region. This agrees with the experimental conclusions presented in [24].

Furthermore, we showed that the number of wind farms within a region does not directly influence the amount of observed smoothing. However, as the number of wind farms increases, the likelihood of observing smoothing increases. Having conducted an in-depth examination of the properties of regional production we examined the relations existing between regional production and available explanatory variables such as single wind farm production measurements and NWP forecasts. We showed that estimating the regional production from a small number of wind farm productions is feasible and that the relation between regional production and single wind farm production is highly linear. We also showed that the relation between single wind farm productions and regional production is non-stationary which also argues in favour of adaptive forecasting models capable of on-line explanatory variable selection. With respect to NWP forecasts we showed that, not surprisingly, wind speed forecasts provide the most useful information on the future regional production. We also illustrated the decreasing explanatory strength of NWP forecasts with increasing prediction horizons. Most importantly, we showed that all NWP model grid points do not provide equal amounts of information on the regional production; some grid points provide consistently better information than others.

From the results of the characterization, we continued by examining the properties of statistical models in the frame of the regional forecasting. We showed that base line forecasting models can be combined in three basic ways, from which an almost infinite number of different model configurations can be obtained. Having defined the three basic configurations we examined to what extent they lead to differences in forecast accuracy and if differences in the base-line models used in these configurations lead to differences in performance.

In order to investigate these questions we defined a new forecasting model, RPC, which we benchmarked on several case studies from the state of the art. This model showed performance levels comparable with those of other state-of-the-art models, with the added advantage that its architecture definition and parameter estimation procedures are very fast. Given its simplicity, this model can be used as a reference model in future studies. Comparing the performance of new models to that of the RPC model will lead to a more significant evaluation of a model's performance than comparing it to naive models such as persistence. Using the RPC and the F-NN models we examined the forecast quality obtained using the three basic modelling approaches. The conclusions from this study are that there is little difference in performance between modelling approaches. The selection of the modelling approach should, in our sense, answer a pragmatic requirement that is: provide the forecasts that are useful to the end-user. Based on the needs of the end-user and the explanatory data available to him (NWP forecasts, measured data, existing single wind farm forecasts) a reasonable combination can be found that provides accurate and useful forecasts. When designing such a tool, the modeller should keep in mind the requirements of the end-user, the constraints of the specific regional forecasting problem and limit, as much as possible, the complexity and the dimensionality of the base model combinations. To further support these guidelines, we showed that the results obtained with the RPC and F-NN model are very similar in terms of performance and no model clearly outperforms the other. There is probably no single base line model better adapted to regional forecasting. Rather, it is the ability of the modeller and his knowledge of his model's and the problem's properties that will lead to differences in performance. The modeller should then use a model whose modelling capacity is sufficient and whose architecture and parameter estimation procedure are well known to him. In this way, the modeller can concentrate on adequately examining the specific properties of the regional forecasting problem that he is faced with.

To further provide modellers with useful insight on how to design regional forecasting models, we investigated the problem of choosing the explanatory variables that will provide the forecasting model with the most useful information on the process under consideration. With this in mind, we proposed a framework to examine the impact that using different explanatory variable subsets can have on regional forecasting model accuracy. We showed that measured power variables provide essential information with respect to the short-term evolution of the regional production, whereas forecast variables, especially wind speed forecasts provided by NWP models, are of prime importance when forecasting the mid and long-term evolution of regional wind power production. We also showed that the amount of noise present in the measured power and forecast wind speed variables can condition the number of different variables of each type that a forecasting model can successfully handle. Since measured power variables are not very noisy, forecasting models can handle a larger number of them. In this way a forecasting model can be provided with a very accurate estimation of the meteorological situation and of the level of production at the time the forecast is being computed. Conversely, forecast wind speed variables are much noisier and, since they often stem from the same model, they present much higher interdependencies. The higher amount of noise implies that a statistical forecasting model can usefully exploit fewer such variables. Also, the dependence between variables means that they are to a certain extent redundant. This redundancy also limits the number of such variables that are useful to a model since increasing their number does not provide the forecasting model with proportionally more information on the future meteorological situation. The results of the simulations showed that relatively few (between 5 and 8) NWP wind speed variables are sufficient to adequately forecast regional production. Using more variables does not necessarily improve forecast performance and can actually lead to lesser forecast accuracy.

From these findings it became apparent that performing variable selection is necessary when designing a regional forecasting model. Not only should the number of explanatory variables be kept within certain limits, but in the case of regional forecasting a modeller is faced with a number of explanatory variables an order of magnitude higher than in the single wind farm forecasting case. To aid the modeller in this task we examined the possibility of applying variable selection procedures to the regional forecasting problem; the particularity of this problem being that the explanatory variables available to the forecaster are all relevant to the process being modelled. From the properties of the regional production revealed by the characterization, and from those observed when examining the influence of explanatory variables on forecast performance, two variable selection heuristics were proposed: the clustering heuristics and a wrapper method. This latter method takes advantage of the transition stability observed between the best variable subsets of different cardinalities. These two heuristics were compared to a state-of-the-art heuristics based on the use of the mutual information measure, the MIFS algorithm. We showed that all three heuristics allow the selection of explanatory variable subsets that lead to good forecasting accuracy. Further the three heuristics present increasing computational cost. The clustering heuristic is the least expensive in this sense, but does not perform as well as the others in terms of forecast accuracy. The next most expensive method is the MIFS algorithm, which presents good performance and reasonable computational cost. Finally, the wrapper approach leads to the best forecasts but requires significantly more computation time. From the results obtained with these heuristics it is clear that variable selection can be advantageously performed in the frame of regional forecasting. Also, the original heuristics proposed in this work provide a first answer to the variable selection need in the frame of regional wind power forecasting.

6.2 Perspectives

The aim of the work presented here was to provide modellers facing a regional forecasting problem with guidelines on relevant problems that must be specifically addressed. We have investigated the different modelling approaches as well as different base models. We have also proposed three variable selection heuristics that can help avoid the curse of dimensionality. However, other aspects of the regional forecasting problem were not investigated.

One such aspect is the possibility of rendering regional forecasting models adaptive in order for them to better cope with the non-stationary nature of regional wind power production. This can seem a straightforward problem given the existence of numerous adaptive forecasting models in the literature, indeed the F-NN model can be rendered adaptive and so can the RPC model. The crux of the problem is that in order to adapt the forecasting models must be capable of computing their past error. This implies that the total regional production be measured and that these measurements be provided to the forecasting models. The total regional production is however seldom available in real, or near real-time. In countries where wind power penetration is high, production data must be collected from very many farms, and the total hourly production is only known with a lag of several weeks. Hence the adaptation of the models will also be lagged. From this situation several questions arise. How does this lag influence the accuracy of the forecasting models? What is the maximum lag that leads to acceptable improvement on forecast accuracy and how much would forecast accuracy be improved by reducing this lag? Answering these questions would allow quantifying the trade off between the amount of lag and the accuracy of the models. This in turn would help decision makers decide if investing to reduce the lag is worthwhile with respect to the economic impact of more accurate forecasts.

Another facet of the regional forecasting problem that would be worth investigating is the possibility of providing probabilistic forecasts of the future regional production. An important amount of research has been undertaken in recent year in order to provide probabilistic forecasts for single wind farm production. Providing the same type of forecasts for regional production could prove useful to utilities wishing to fully exploit wind power forecasts in their management processes, i.e unit commitment, economic dispatch, etc. From this point of view the single wind farm error characterization performed in [36] could be extended to the regional forecasting problem. From this, probabilistic regional forecasting models such as the one described in [145] could be implemented. Given the statistical error smoothing that occurs in regional forecasts, the forecast distributions that could be obtained might well be proportionally sharper than those observed in the frame of single wind farm forecasts. This in turn could lead stochastic power system scheduling models, comparable in principle to that described in [147], to be used in the frame of power system management in order to minimize the added cost incurred by networks facing important wind power penetration.

In the same perspective the potential improvement that could be derived from the consideration of ensemble weather forecasts in regional wind power forecasting should be investigated. The use of such NWP forecasts has been examined for single wind farm forecasts and more specifically in order to provide enhanced probabilistic forecasts. This possibility exists with respect to regional forecasting, especially since it has often been observed that the use of the average of the members of an ensemble forecast leads to more accurate single wind farm forecasts than using the control member [148]. In such an examination however, care should be taken to intelligently extract the added information that the ensembles provide without exposing the regional forecasting model to high input dimensions.

From a more general perspective the investigation framework described in this thesis should be extended to a higher number of case studies involving larger and more varied regions. Regions such as Spain and France would be of interest given the diversity of their differing meteorological situations and the variety of their terrain complexities. In this way the positive effect of lower correlation between different wind regimes could be studied in more detail, the characteristics of regional forecasting described in the present work could be further documented and new characteristics detected. Such a study could also lead to a forecasting competition comparable to that performed for single wind farm forecasts in the frame of the European research project Anemos. Such a study would yield results that could undergo an in-depth analysis so that further knowledge on the regional forecasting problem can be gained from the model design perspective.

Further, another research perspective that merits investigation is the possibility of integrating meteorological measurements in the regional forecasting process. As shown in [67] the bias of wind speed forecasts can be reduced when using wind speed measurements to filter the raw NWP forecasts. This possibility should be further investigated in itself and in the frame of regional forecasting. As shown in the present work, a rule of thumb for reference wind farm selection is to select wind farms that a priori provide meteorological information over the whole region. From this it might be of interest to examine the possibility of integrating wind speed measurements taken within, and outside, the region in order to correct the NWP forecast values used as inputs to the regional power forecasting models. This approach could constitute a first, statistical, step toward the improvement of the NWP data. A second, possibly more complex step, would involve the development of dedicated models capable of exploiting meteorological measurements in order to improve the accuracy of wind power forecasts. One possibility would be to combine physical and statistical modelling so that through fast data assimilation and short computation times, conventional NWP forecasts can be improved, at least for the short-term horizons. Such an approach could also possibly provide more reliable forecasts of extreme situations, such as severe storms, where massive regional wind power production variations are likely to be observed. Such forecasts, given even within three hours of the event, could prove decisive to the economic feasibility of large-scale wind power integration in interconnected networks.

APPENDIX A

Complementary Characterization Results

A.1 Smoothing and aggregation results for the Irish case study

In the following, the results concerning the smoothing factor S with respect to wind farm properties and the number of aggregated wind farms are presented. Are also presented the results concerning the correlation between the aggregated production of subsets of wind farms and the total production. The following figures are presented:

- Smoothing factor versus virtual region size (Figure A.1).
- Smoothing factor versus the average distance between wind farms in the combinations (Figure A.2).
- Relation between region size and mean distance between wind farms on the smoothing factor (Figure A.3).
- Smoothing factor, and mean smoothing factor versus the number of wind farms in each combination (Figure A.4).
- Correlation between the aggregated production of the 11 Irish wind farms and the aggregated wind farm combination production, versus the cardinality of the wind farm combinations (Figure A.5).

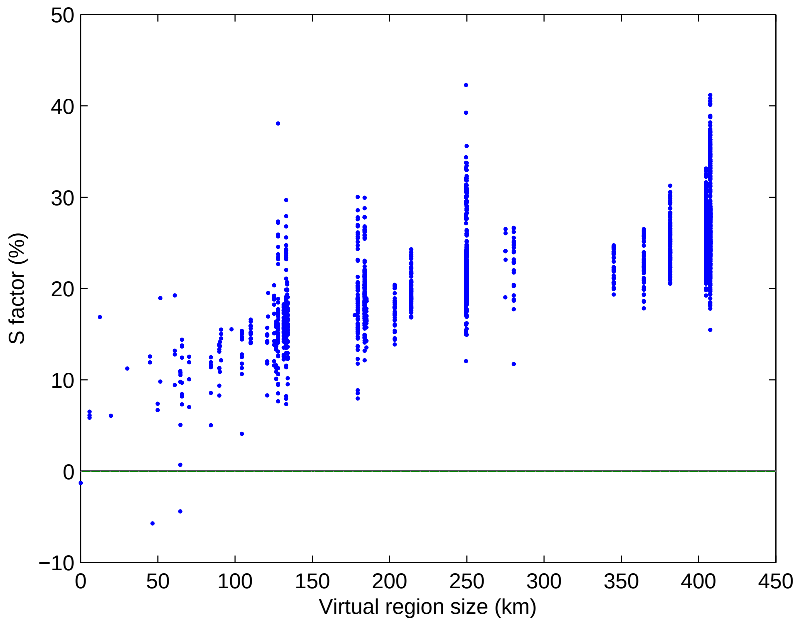


FIGURE A.1: *Smoothing factor versus virtual region size for the Irish case.*

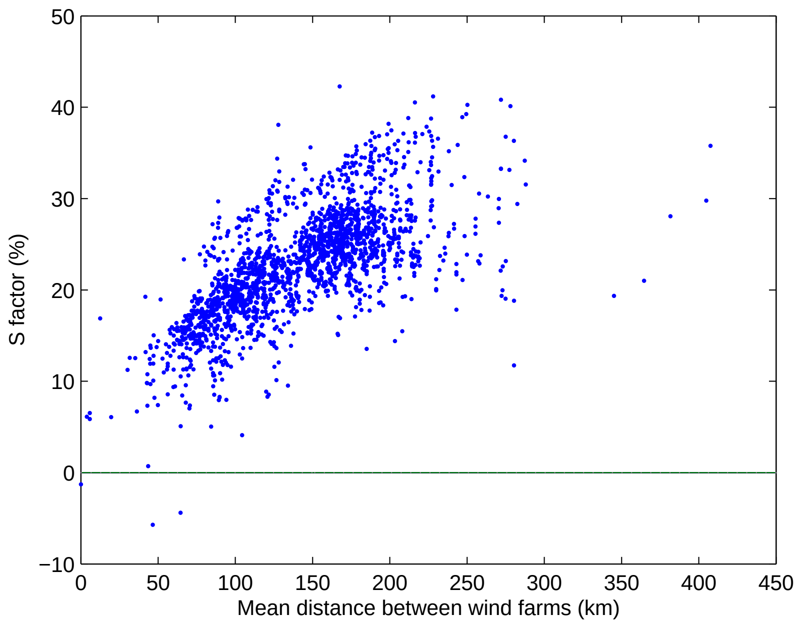


FIGURE A.2: *Smoothing factor versus the average distance between wind farms in the combinations, for the Irish case.*

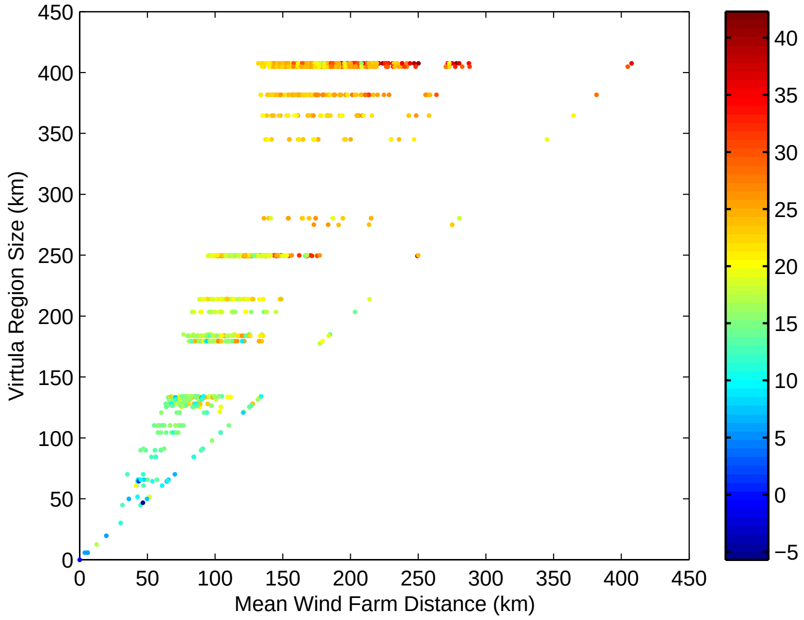


FIGURE A.3: *Relation between region size and mean distance between wind farms on the smoothing factor, for the Irish case. The colour bar on the right of the figure indicated the smoothing factor value for each wind farm combination.*

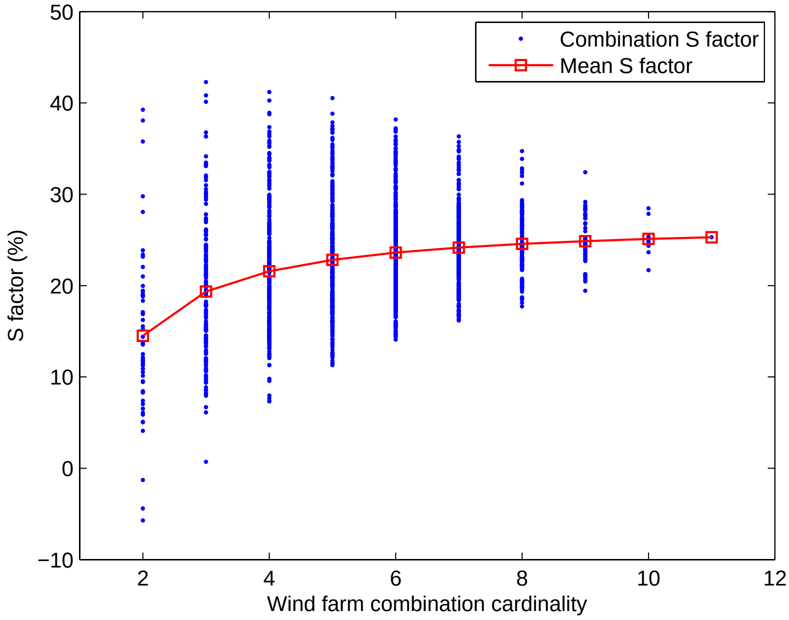


FIGURE A.4: *Smoothing factor, and mean smoothing factor versus the number of wind farms in each combination, for the Irish case.*

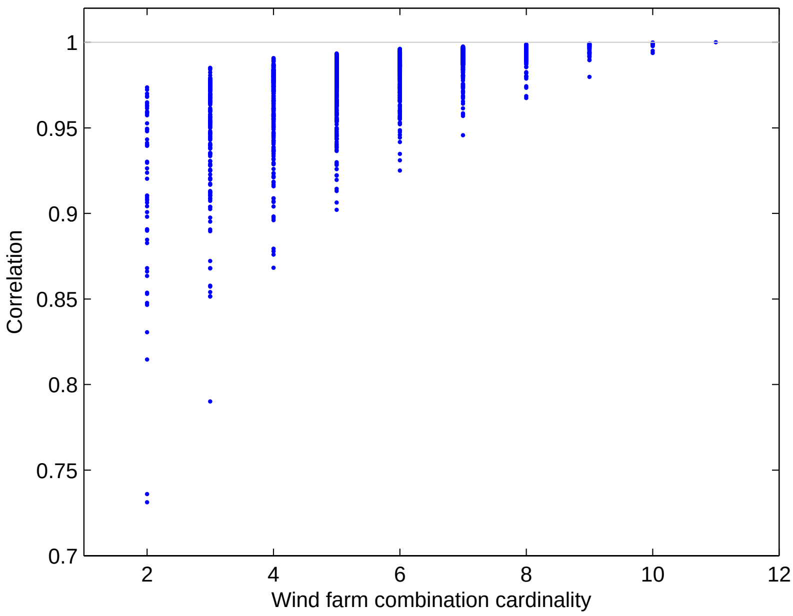


FIGURE A.5: *Correlation between the aggregated production of the 11 Irish wind farms and the aggregated wind farm combination production, versus the cardinality of the wind farm combinations*

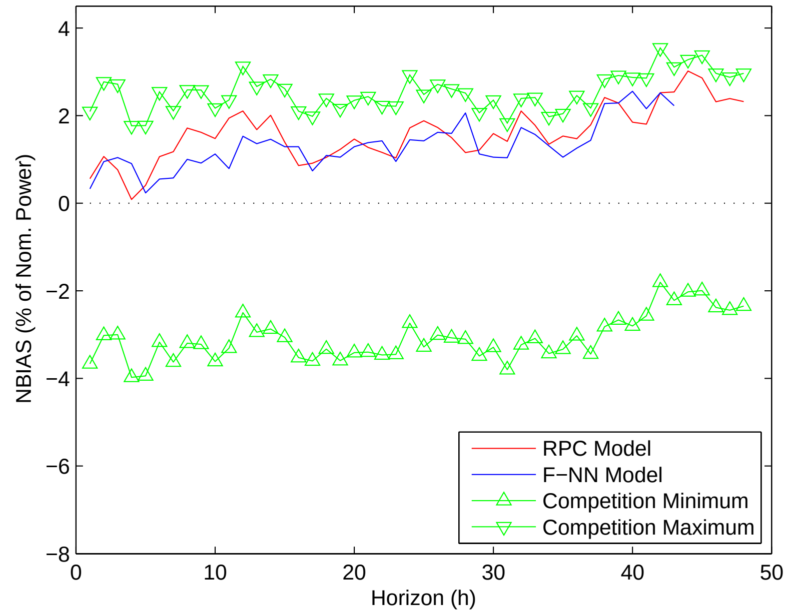
APPENDIX B

Complementary Regional Forecasting Results

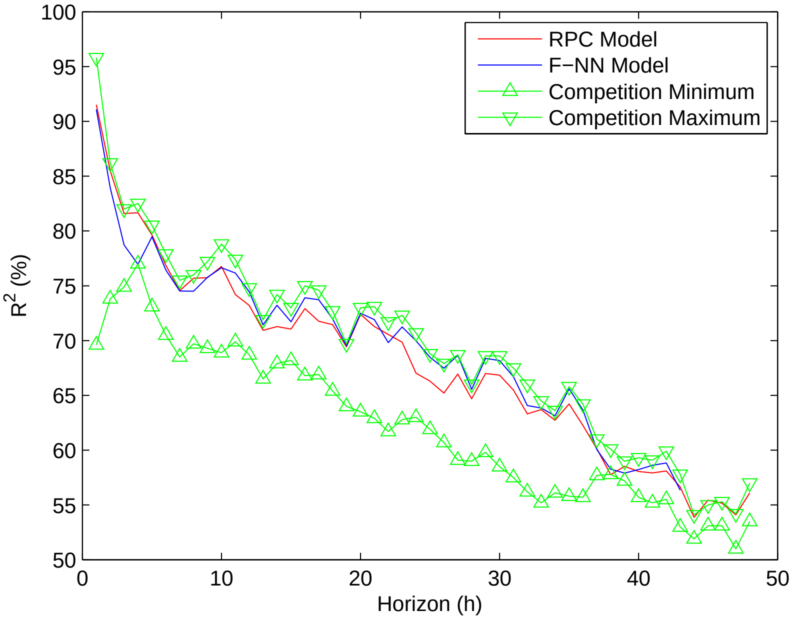
B.1 Validation of the RPC model for single wind farm forecasting

In this section the NBIAS and R^2 values found using the F-NN and RPC models are presented for the single wind farm forecasts of Tunø Knob, Golagh and Alaiz. The values are compared to the maximum and minimum values obtained by the models that took part in the Anemos Competition. These results are part of the validation of the RPC model described in subsection 4.5.1. The following figures are presented:

- NBIAS versus forecast horizon for Tunø Knob (Figure B.1(a)).
- R^2 versus forecast horizon for Tunø Knob (Figure B.1(b)).
- NBIAS versus forecast horizon for Golagh (Figure B.2(a)).
- R^2 versus forecast horizon for Golagh (Figure B.2(b)).
- NBIAS versus forecast horizon for Alaiz (Figure B.3(a)).
- R^2 versus forecast horizon for Alaiz (Figure B.3(b)).

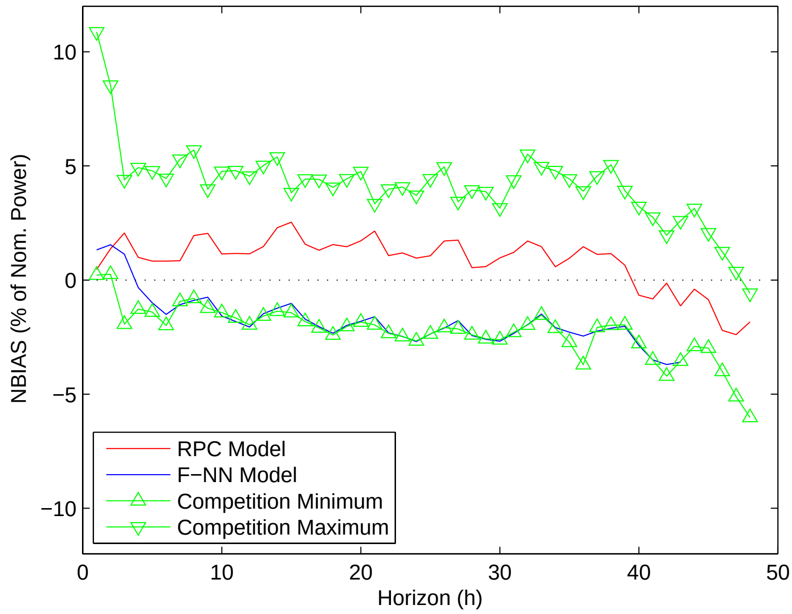


(a) Tunø Knob BIAS

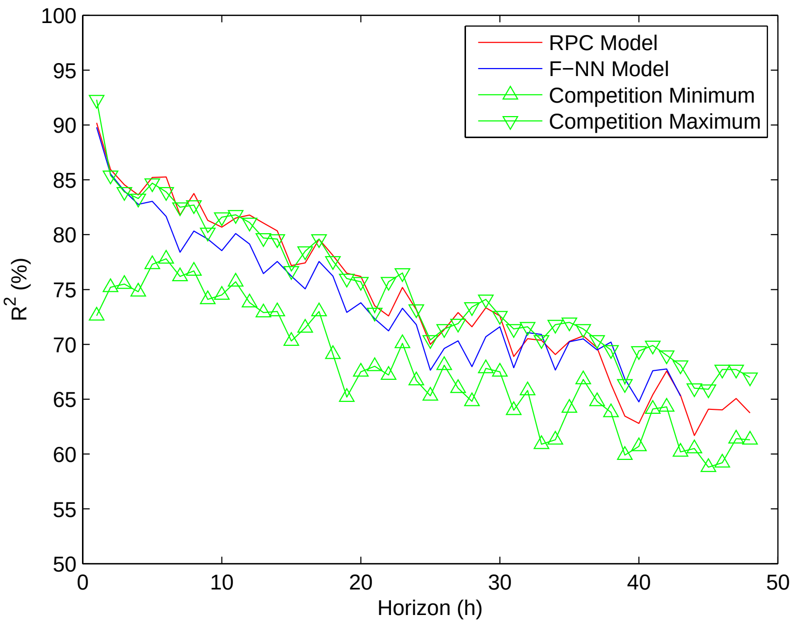


(b) Tunø Knob $R2$

FIGURE B.1: *NBIAS and R^2 as a function of forecast horizon of the RPC and F-NN model for Tunø Knob wind farm. The green curves represent the minimum and maximum values obtained by the state-of-the-art models that participated in the Anemos Competition.*

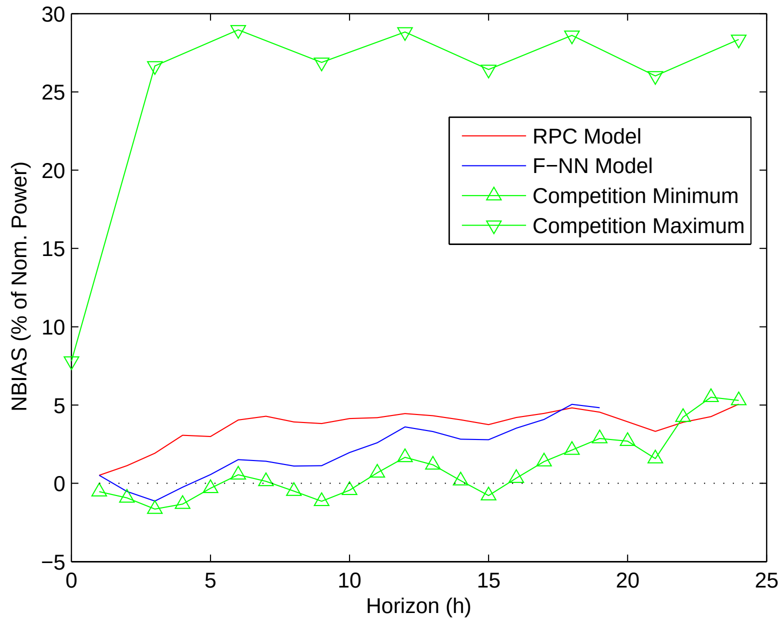


(a) Golagh NBIAS

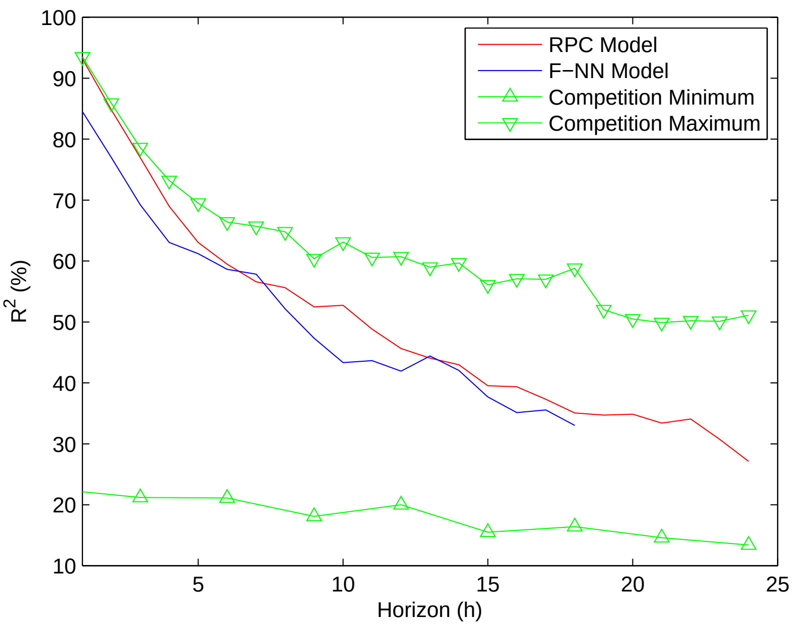


(b) Golagh R^2

FIGURE B.2: *NBIAS and R^2 as a function of forecast horizon of the RPC and F-NN model for Golagh wind farm. The green curves represent the minimum and maximum values obtained by the state-of-the-art models that participated in the Anemos Competition.*



(a) Alaiz NBIAS



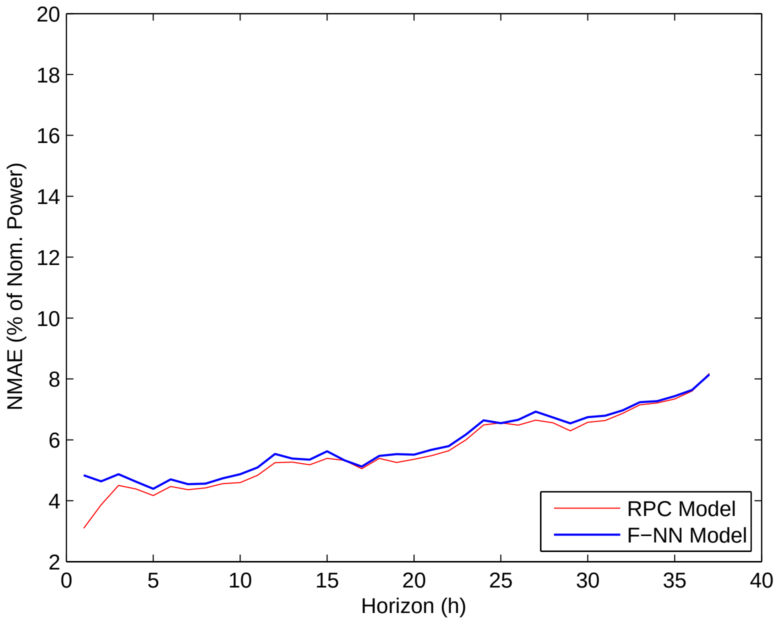
(b) Alaiz R^2

FIGURE B.3: *NBIAS and R^2 as a function of forecast horizon of the RPC and F-NN model for Alaiz wind farm. The green curves represent the minimum and maximum values obtained by the state-of-the-art models that participated in the Anemos Competition.*

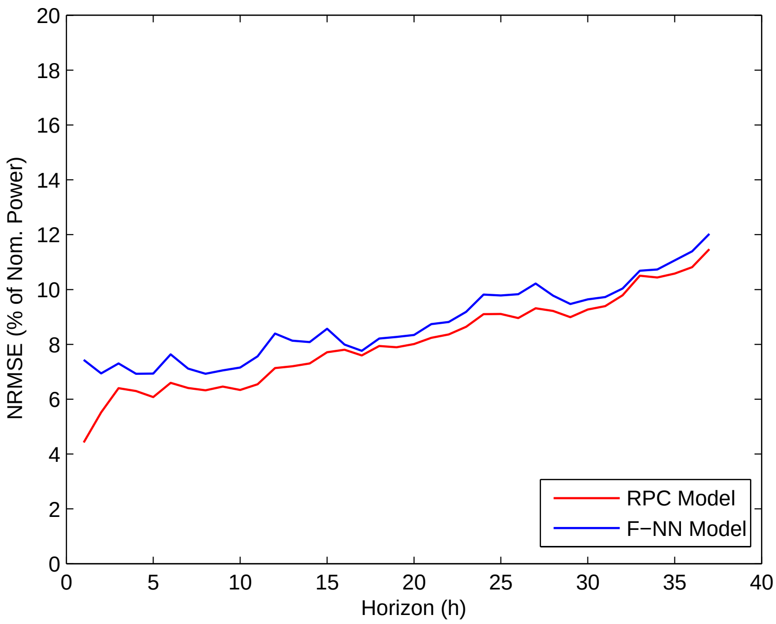
B.2 Direct Approach for the Danish Case

In this section the forecast evaluation criteria found, using the F-NN and RPC models, for the Direct Upscaling approach (described in section 4.5.2) for the case of Denmark are presented. The criteria are presented in the following figures:

- NMAE versus forecast horizon using the RPC and F-NN models (Figure B.4(a)).
- NRMSE versus forecast horizon using the RPC and F-NN models (Figure B.4(b)).
- NBIAS versus forecast horizon using the RPC and F-NN models (Figure B.5(a)).
- Skewness versus forecast horizon using the RPC and F-NN models (Figure B.5(b)).
- NSDE versus forecast horizon using the RPC and F-NN models (Figure B.6(a)).
- Kurtosis versus forecast horizon using the RPC and F-NN models (Figure B.6(b)).
- R^2 versus forecast horizon using the RPC and F-NN models (Figure B.7(a)).
- NRMSE improvement over OL-Persistence versus forecast horizon using the RPC and F-NN models (Figure B.7).

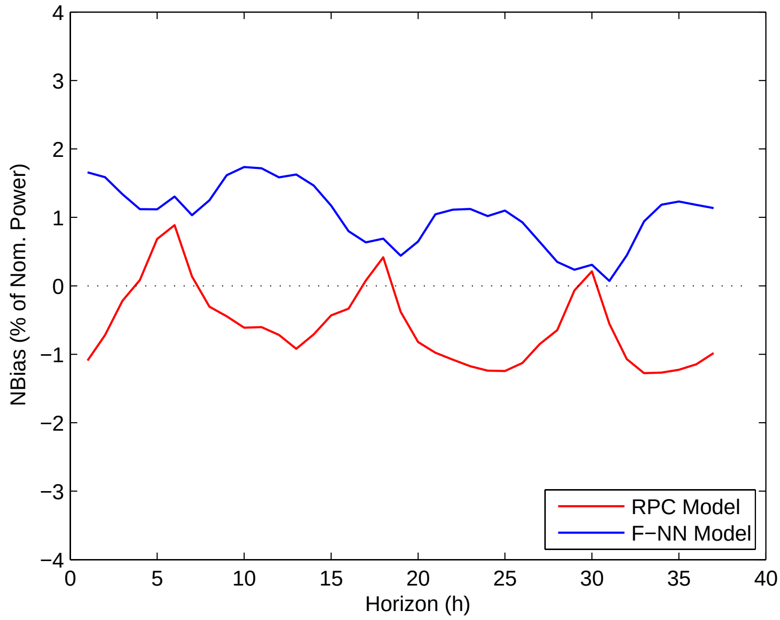


(a) NMAE

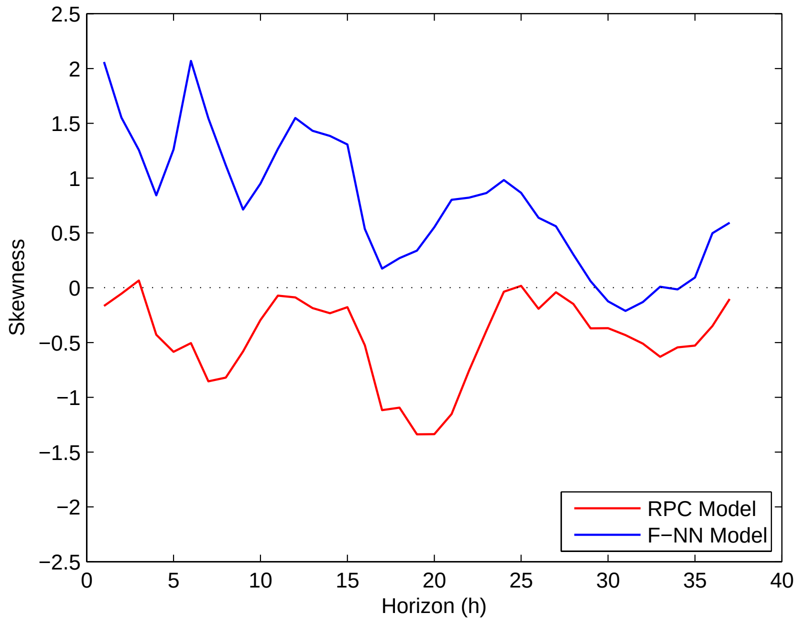


(b) NRMSE

FIGURE B.4: *NMAE and NRMSE of the RPC and F-NN models for the direct upscaling approach for the Danish case. The input data used by both models consists of the average NWP forecasts and on-line power measurements for all wind farms.*

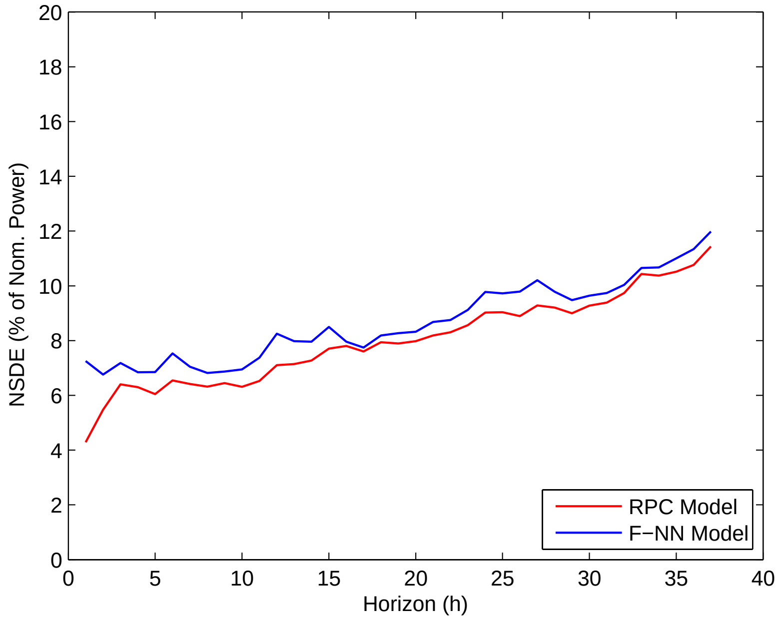


(a) NBIAS

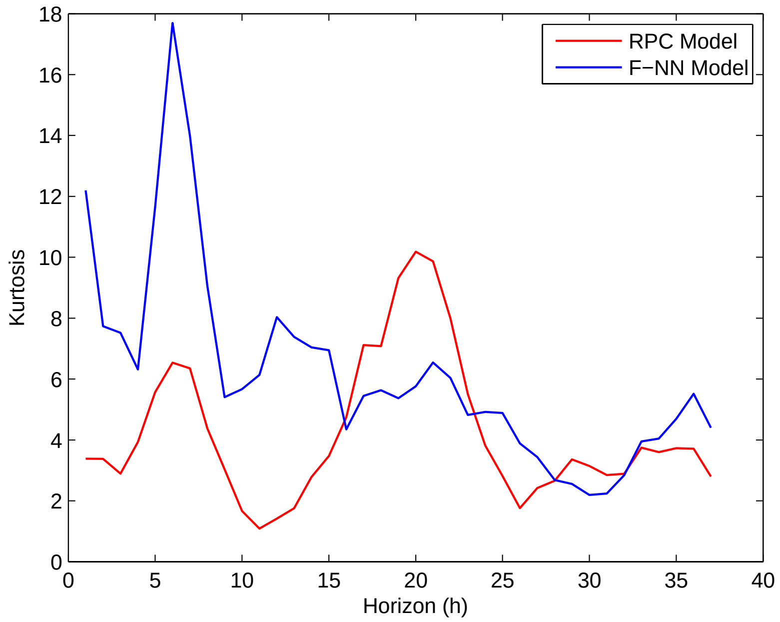


(b) Skewness

FIGURE B.5: *Normalized Bias and Skewness of the RPC and F-NN models for the direct upscaling approach for the Danish case.*

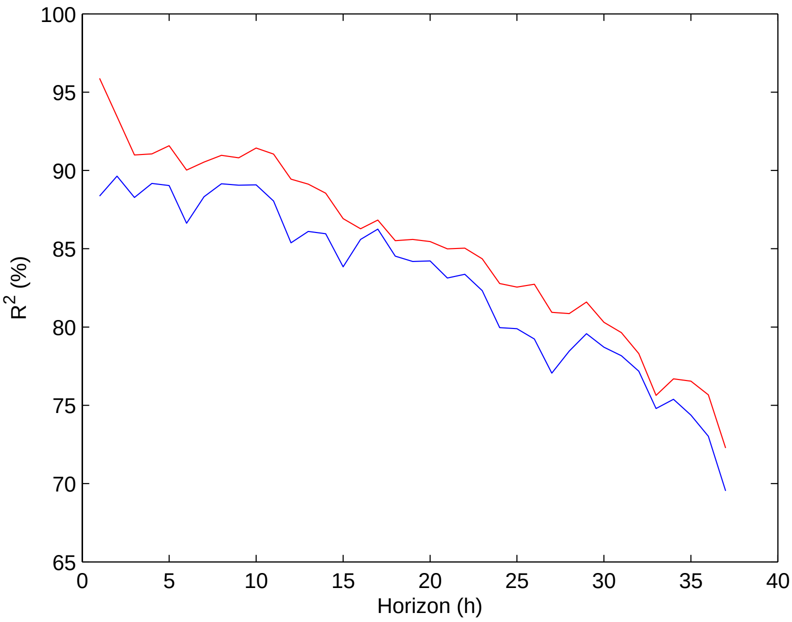


(a) NSDE

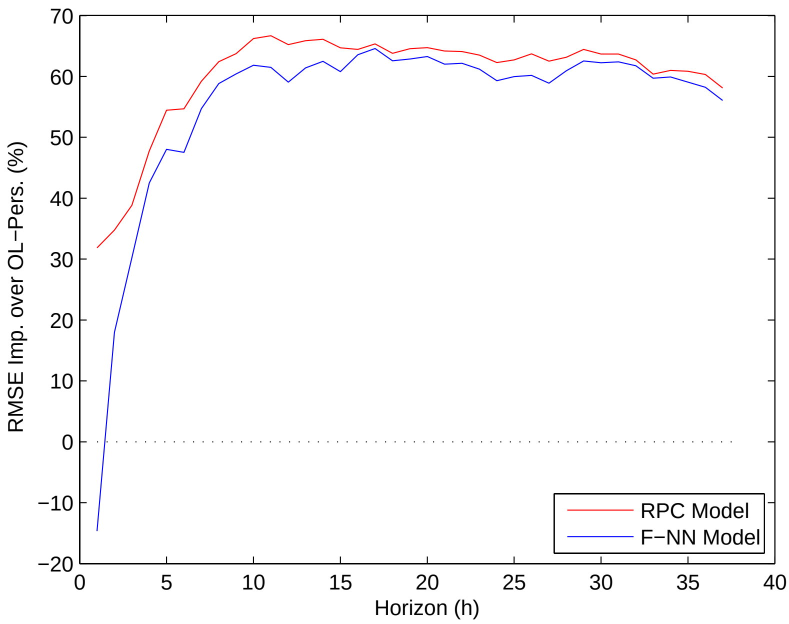


(b) Kurtosis

FIGURE B.6: *NSDE and Kurtosis of the RPC and F-NN models for the direct upscaling approach for the Danish case.*



(a) R^2



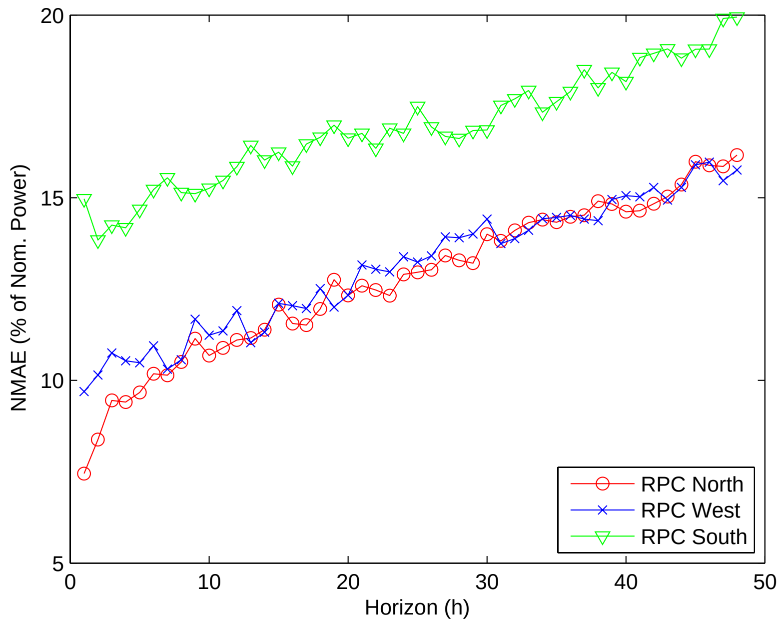
(b) Improvement over OL-Persistence

FIGURE B.7: *R^2 and NRMSE improvement over OL-Persistence of the RPC and F-NN models for the direct upscaling approach for the Danish case.*

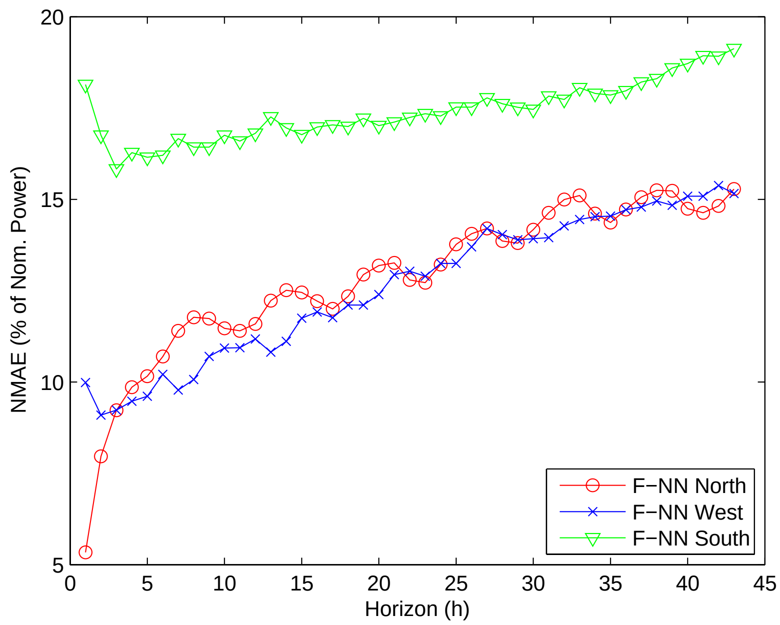
B.3 Cluster Approach for the Irish Case

In this section the forecast evaluation criteria, computed on the individual cluster forecasts of the Cluster approach described in section 4.5.2, are presented. The criteria are computed for both the RPC and F-NN models. The following figures are presented:

- NMAE versus forecast horizon computed for the RPC model (Figure B.8(a)).
- NMAE versus forecast horizon computed for the F-NN model (Figure B.8(b)).
- NRMSE versus forecast horizon computed for the RPC model (Figure B.9(a)).
- NRMSE versus forecast horizon computed for the F-NN model (Figure B.9(b)).
- NBIAS versus forecast horizon computed for the RPC model (Figure B.10(a)).
- NBIAS versus forecast horizon computed for the F-NN model (Figure B.10(b)).
- Skewness versus forecast horizon computed for the RPC model (Figure B.11(a)).
- Skewness versus forecast horizon computed for the F-NN model (Figure B.11(b)).
- NSDE versus forecast horizon computed for the RPC model (Figure B.12(a)).
- NSDE versus forecast horizon computed for the F-NN model (Figure B.12(b)).
- Kurtosis versus forecast horizon computed for the RPC model (Figure B.13(a)).
- Kurtosis versus forecast horizon computed for the F-NN model (Figure B.13(b)).
- R^2 versus forecast horizon computed for the RPC model (Figure B.14(a)).
- R^2 versus forecast horizon computed for the F-NN model (Figure B.14(b)).
- NRMSE improvement over Persistence versus forecast horizon computed for the RPC model (Figure B.15(a)).
- NRMSE improvement over Persistence versus forecast horizon computed for the F-NN model (Figure B.15(b)).

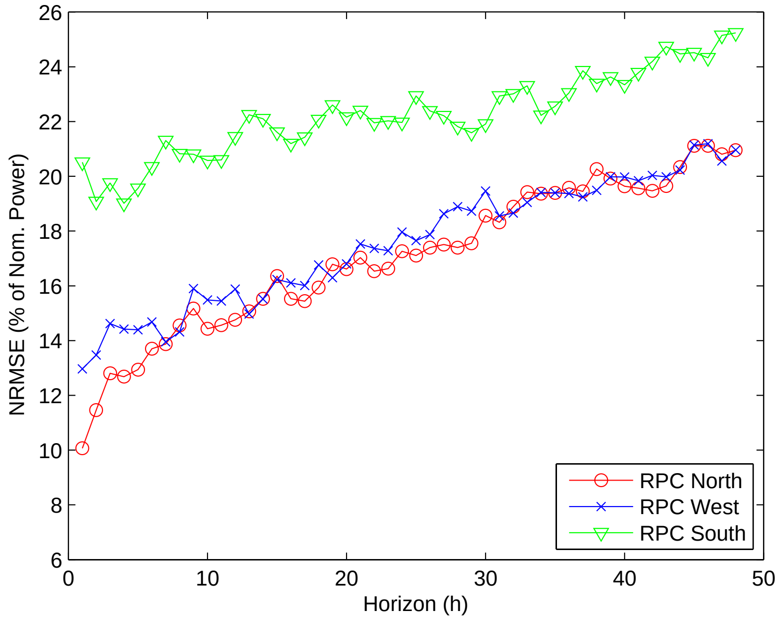


(a) RPC

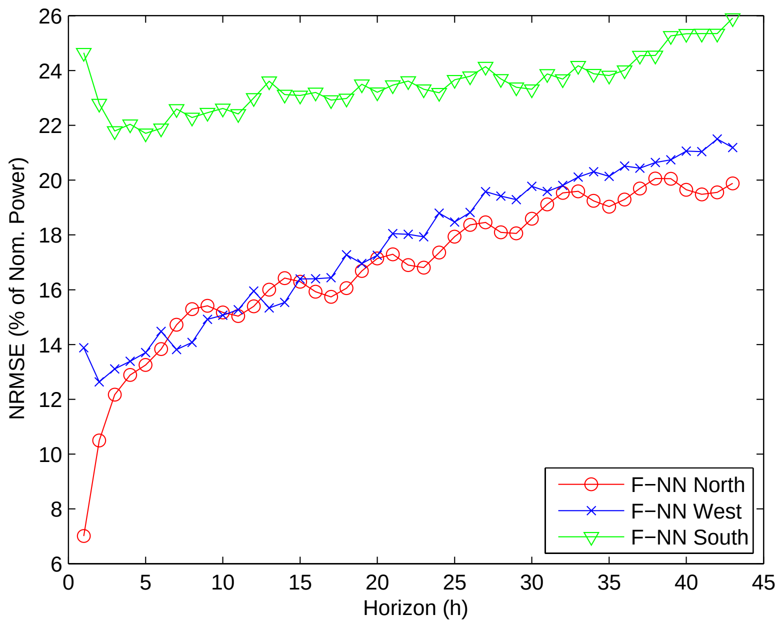


(b) F-NN

FIGURE B.8: *NMAE obtained, using the RPC and F-NN models, for the three cluster forecasts in the Irish case.*

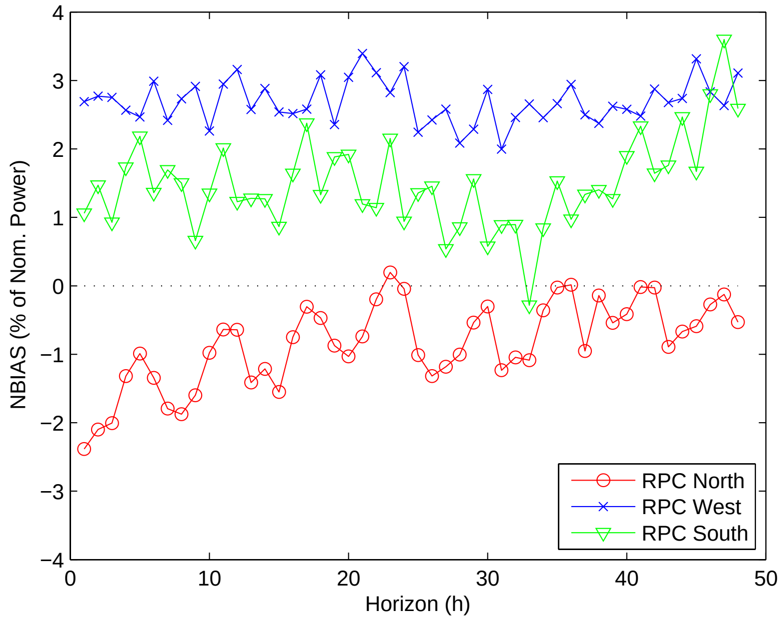


(a) RPC

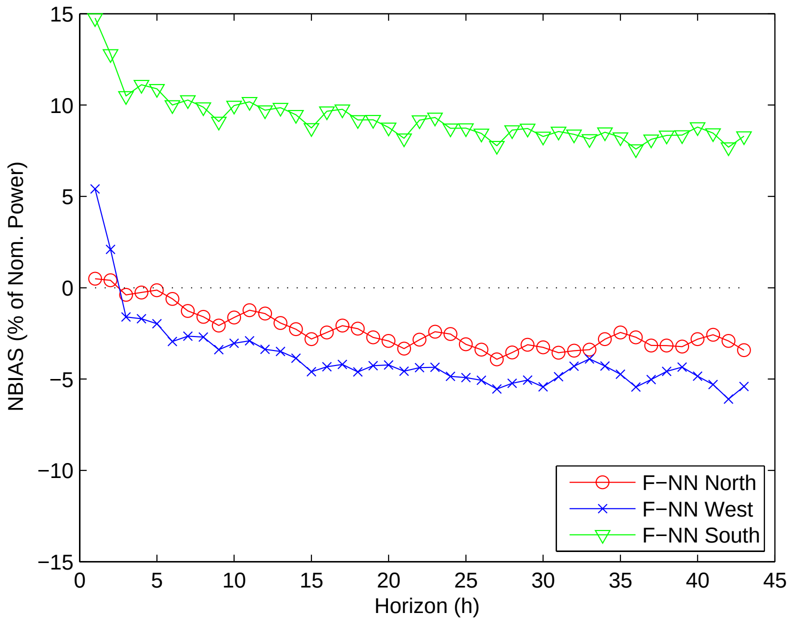


(b) F-NN

FIGURE B.9: *NRMSE obtained, using the RPC and F-NN models, for the three cluster forecasts in the Irish case.*

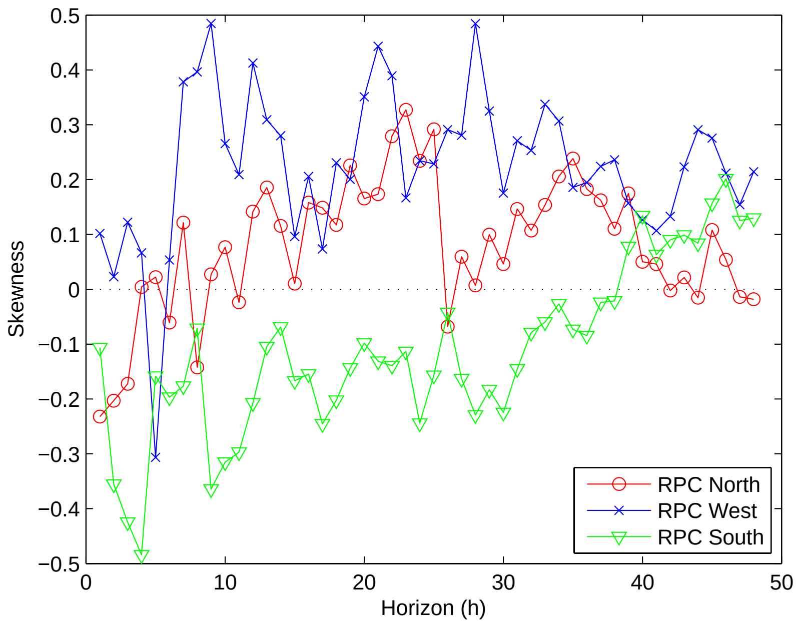


(a) RPC

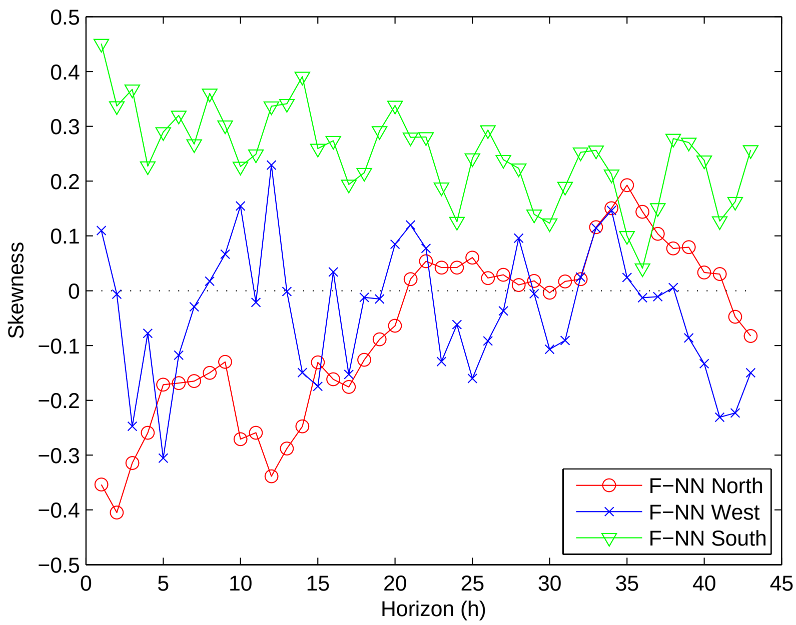


(b) F-NN

FIGURE B.10: *NBIAS obtained, using the RPC and F-NN models, for the three cluster forecasts in the Irish case.*

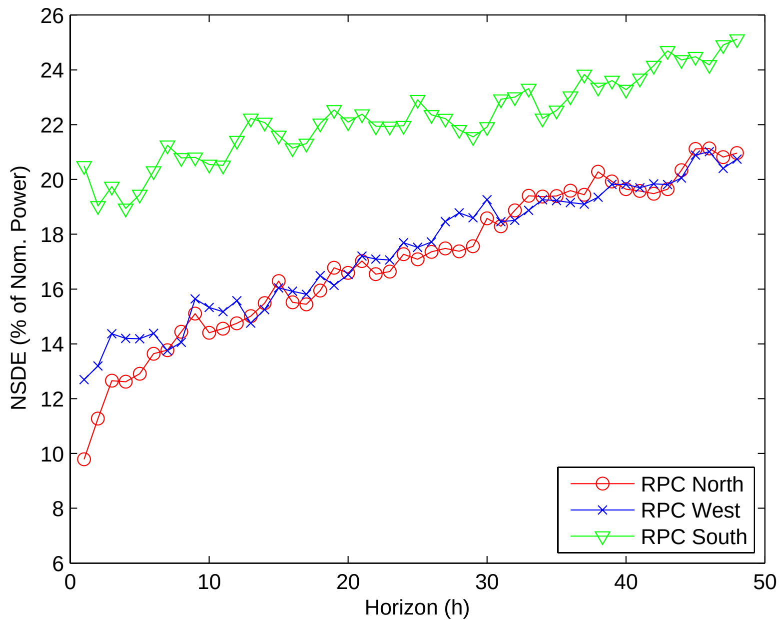


(a) RPC

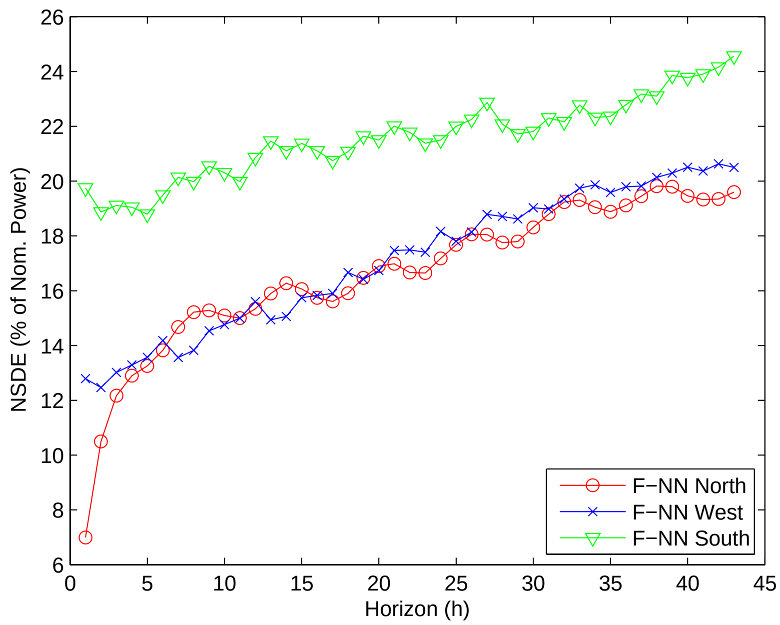


(b) F-NN

FIGURE B.11: *Skewness obtained, using the RPC and F-NN models, for the three cluster forecasts in the Irish case.*

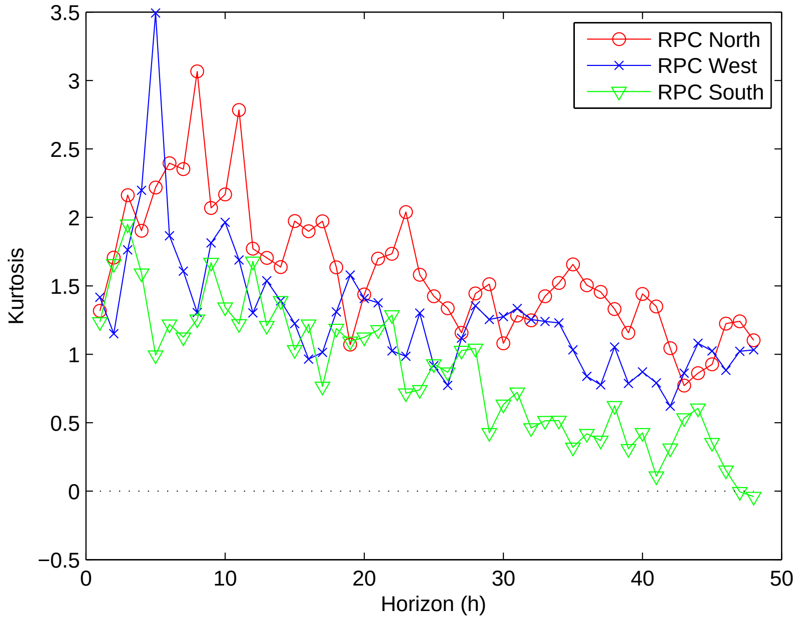


(a) RPC

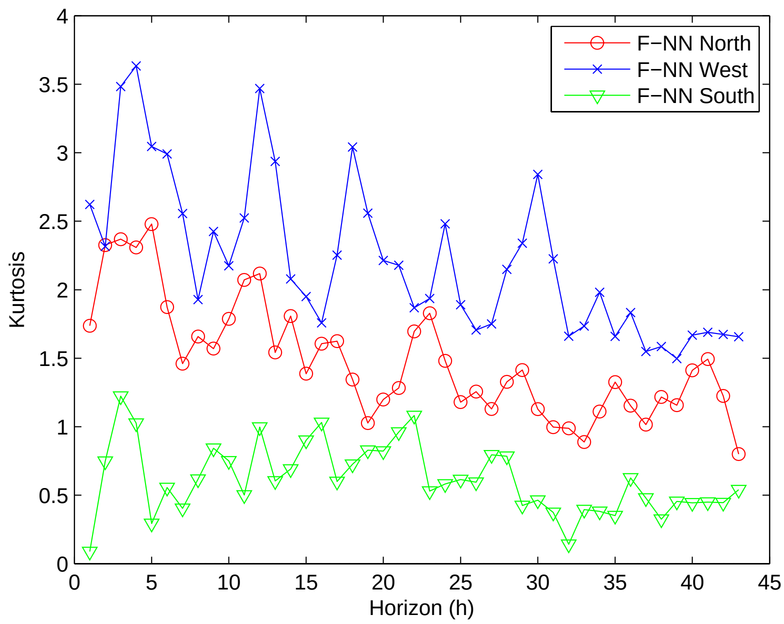


(b) F-NN

FIGURE B.12: *NSDE obtained, using the RPC and F-NN models, for the three cluster forecasts in the Irish case.*

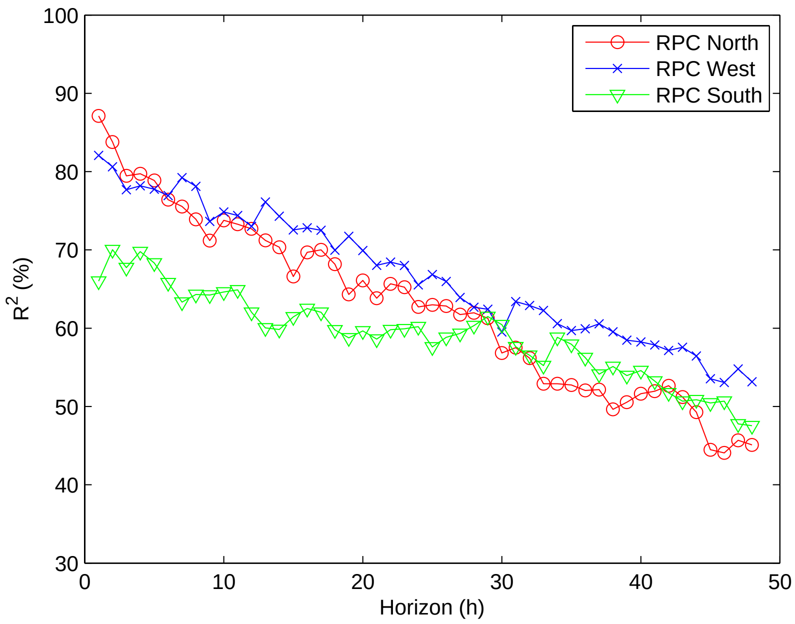


(a) RPC

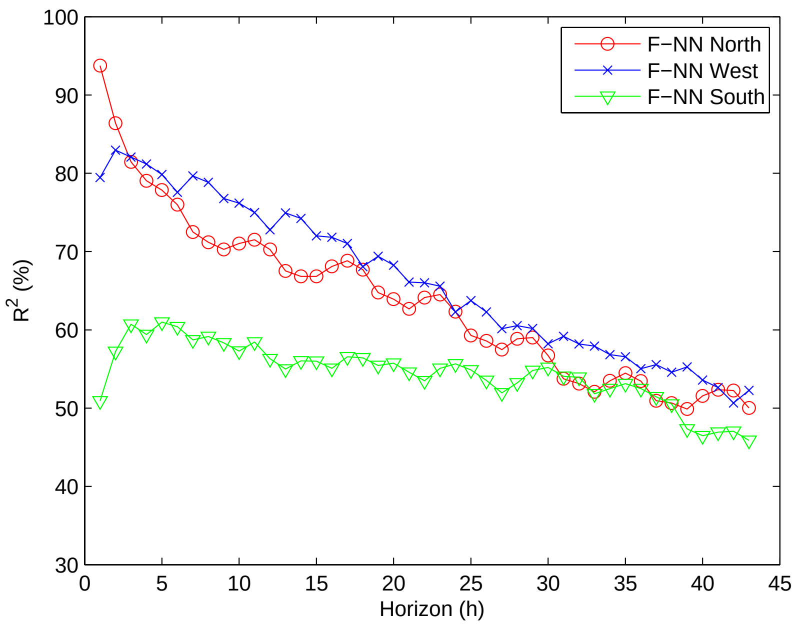


(b) F-NN

FIGURE B.13: *Kurtosis obtained, using the RPC and F-NN models, for the three cluster forecasts in the Irish case.*

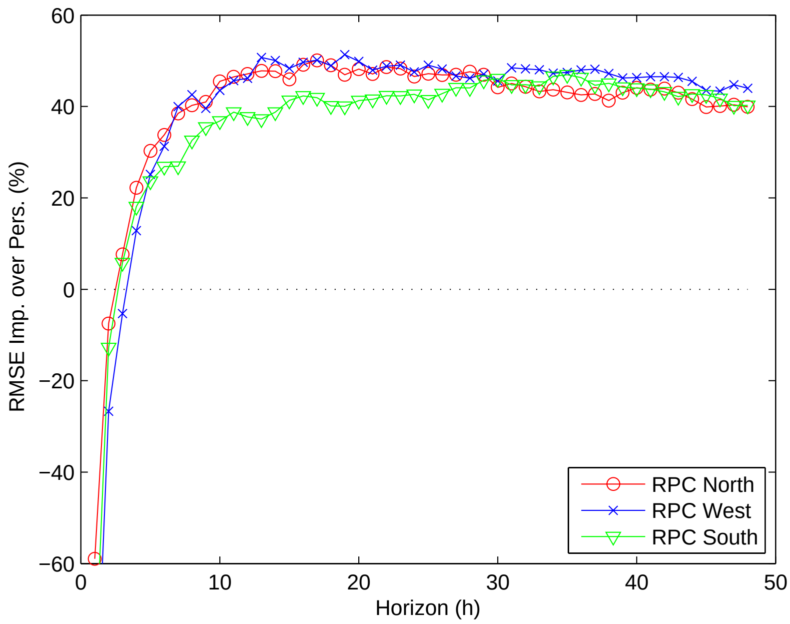


(a) RPC

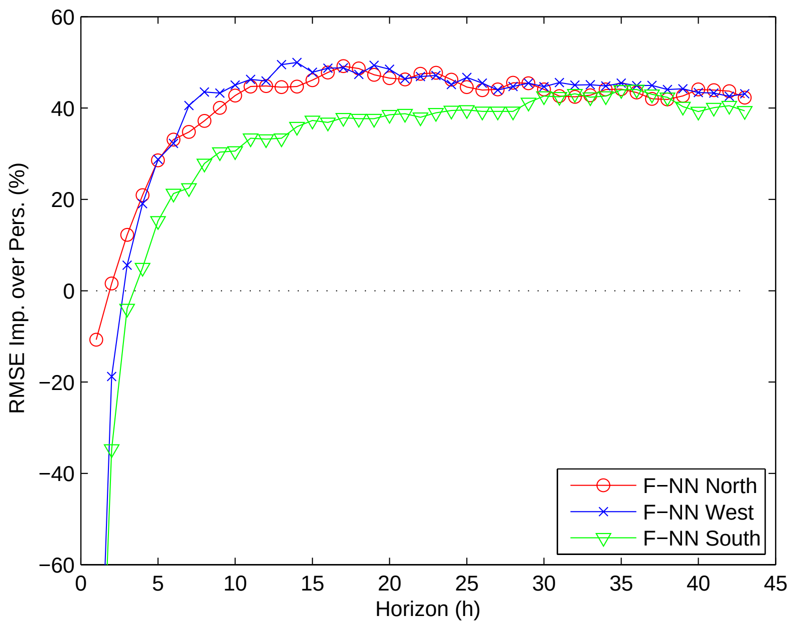


(b) F-NN

FIGURE B.14: *R^2 obtained, using the RPC and F-NN models, for the three cluster forecasts in the Irish case.*



(a) RPC



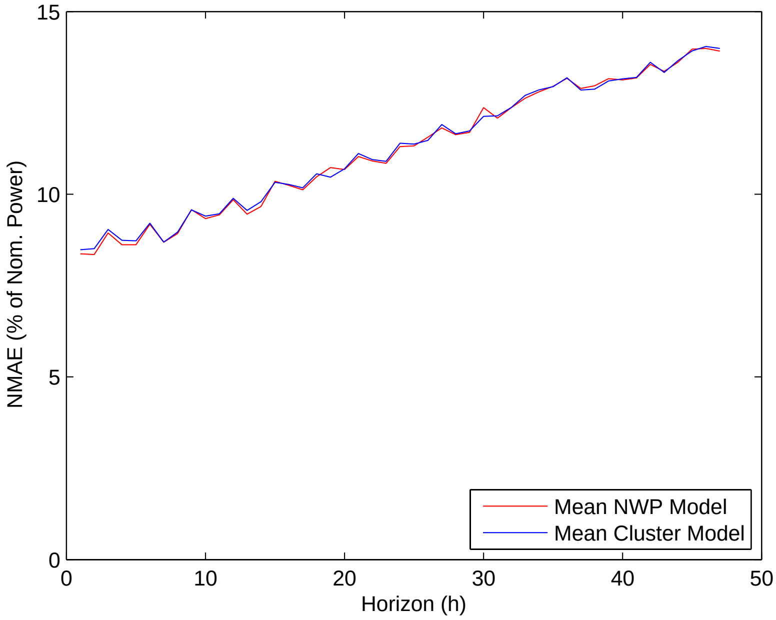
(b) F-NN

FIGURE B.15: *NRMSE improvement over Persistence obtained, using the RPC and F-NN models, for the three cluster forecasts in the Irish case.*

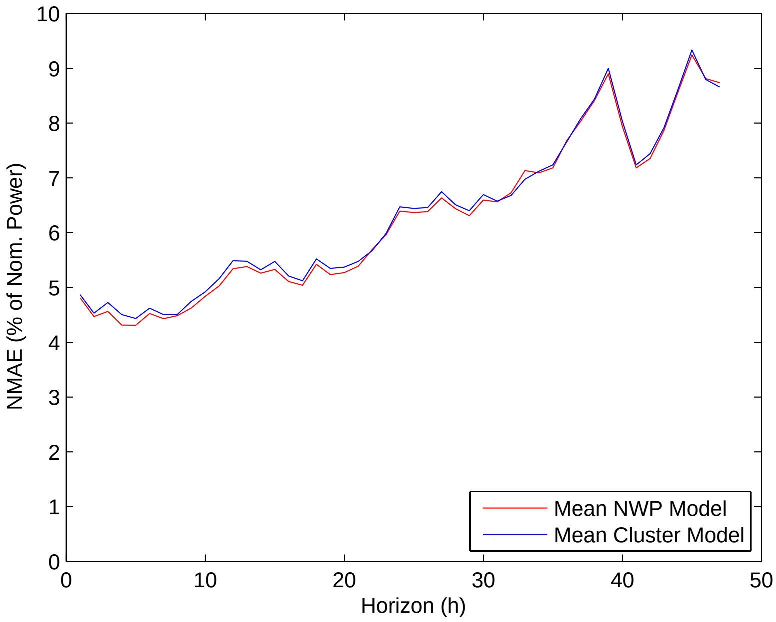
B.4 Comparison of modelling approaches

In this section the forecast evaluation criteria found, using the RPC models for the comparison of two regional forecasting approaches (described in section 4.5.2) are presented. The results are for the cases of Ireland and Denmark using the mean NWP and mean cluster approaches. The following figures are presented:

- NMAE versus forecast horizon for the case of Ireland (Figure B.16(a)).
- NMAE versus forecast horizon for the case of Denmark (Figure B.16(b)).
- NRMSE versus forecast horizon for the case of Ireland (Figure B.17(a)).
- NRMSE versus forecast horizon for the case of Denmark (Figure B.17(b)).
- NBIAS versus forecast horizon for the case of Ireland (Figure B.18(a)).
- NBIAS versus forecast horizon for the case of Denmark (Figure B.18(b)).
- Skewness versus forecast horizon for the case of Ireland (Figure B.19(a)).
- Skewness versus forecast horizon for the case of Denmark (Figure B.19(b)).
- NSDE versus forecast horizon for the case of Ireland (Figure B.20(a)).
- NSDE versus forecast horizon for the case of Denmark (Figure B.20(b)).
- Kurtosis versus forecast horizon for the case of Ireland (Figure B.21(a)).
- Kurtosis versus forecast horizon for the case of Denmark (Figure B.21(b)).
- R^2 versus forecast horizon for the case of Ireland (Figure B.22(a)).
- R^2 versus forecast horizon for the case of Denmark (Figure B.22(b)).
- NRMSE improvement over Persistence versus forecast horizon for the case of Ireland (Figure B.23(a)).
- NRMSE improvement over Persistence versus forecast horizon for the case of Denmark (Figure B.23(b)).

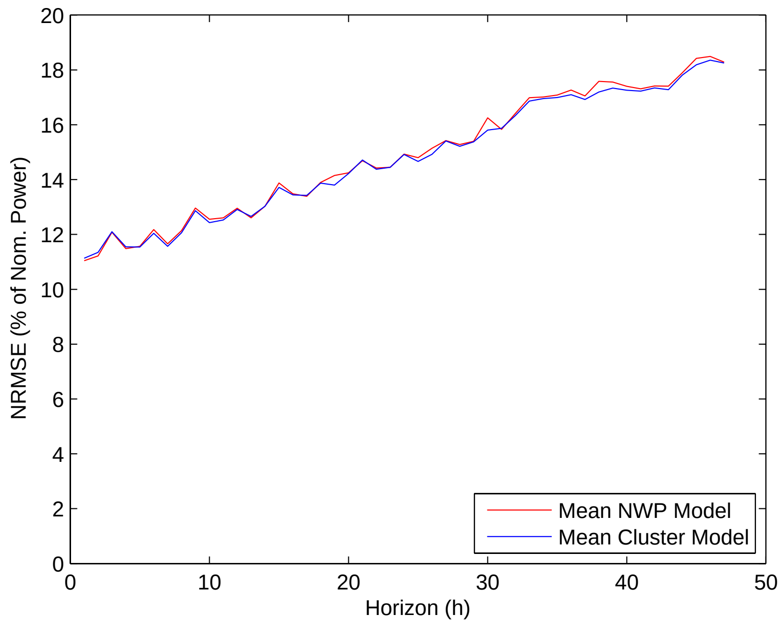


(a) Ireland

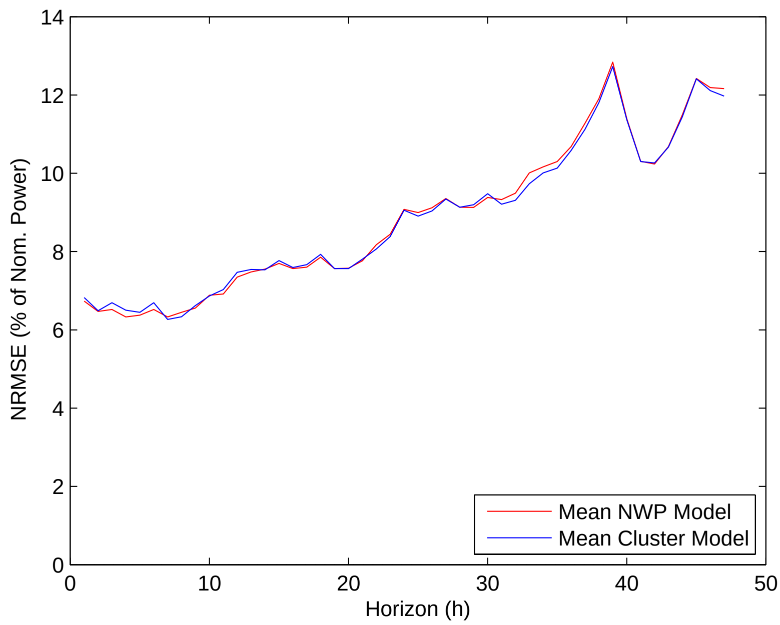


(b) Denmark

FIGURE B.16: *NMAE vs. forecast horizon for the case of Ireland and Denmark using the mean meteorological approach and the mean cluster approach*

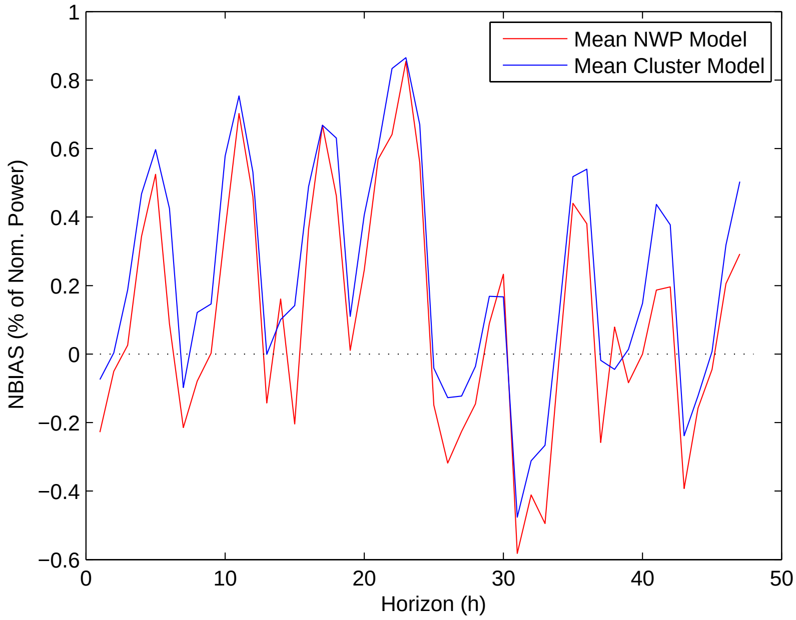


(a) Ireland

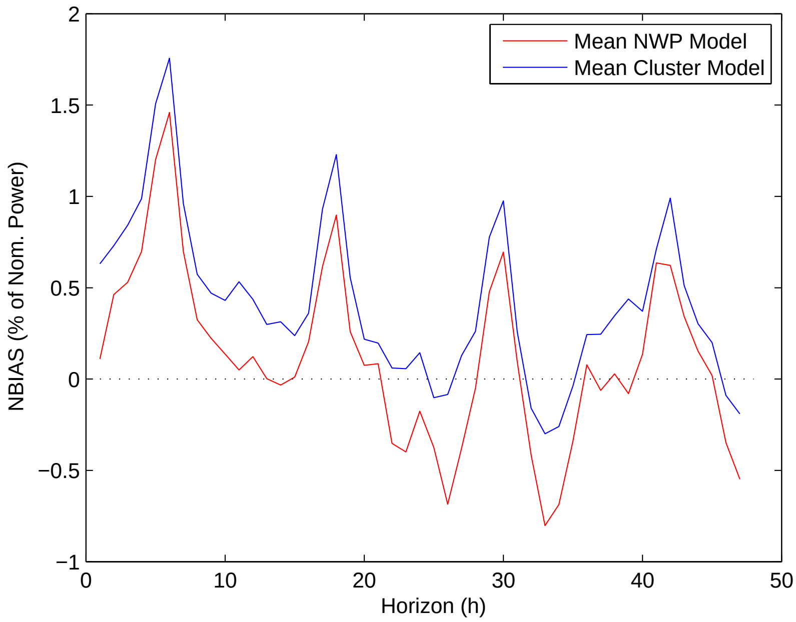


(b) Denmark

FIGURE B.17: *NRMSE vs. forecast horizon for the case of Ireland and Denmark using the mean meteorological approach and the mean cluster approach*

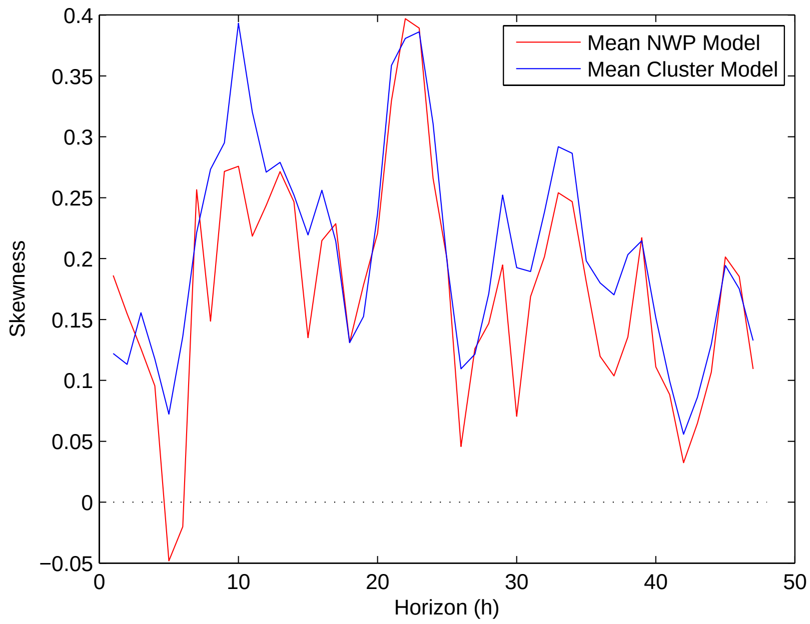


(a) Ireland

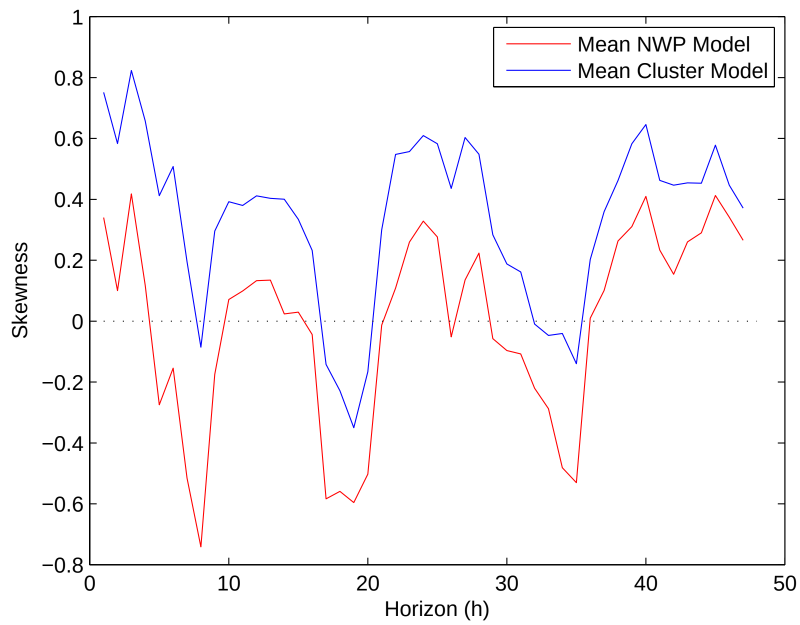


(b) Denmark

FIGURE B.18: *NBIAS vs. forecast horizon for the case of Ireland and Denmark using the mean meteorological approach and the mean cluster approach*

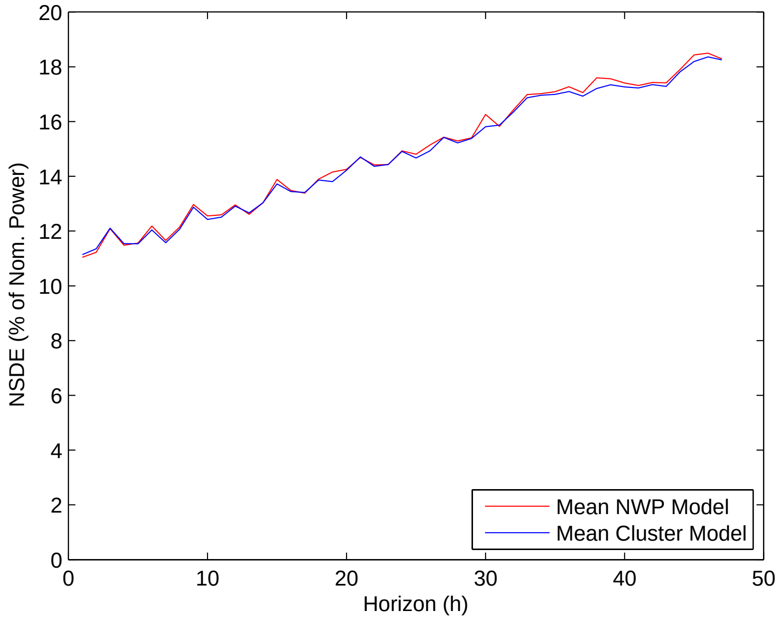


(a) Ireland

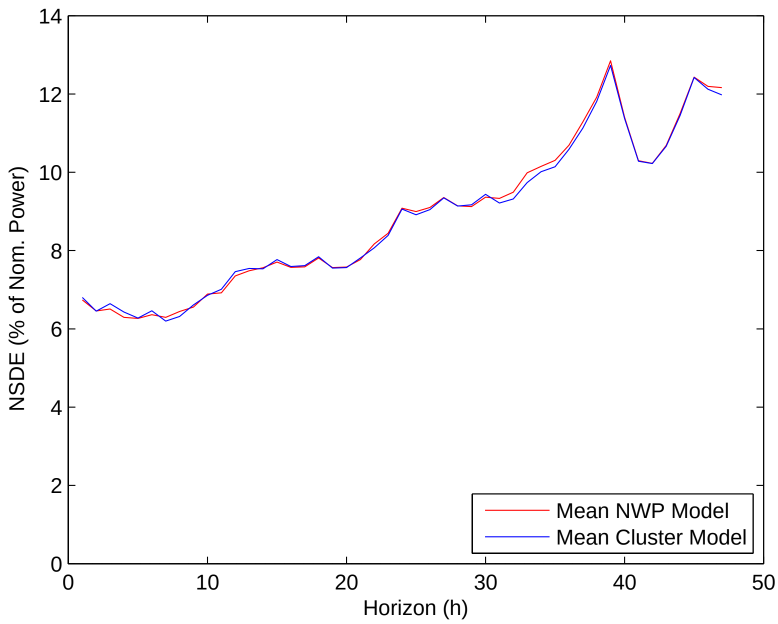


(b) Denmark

FIGURE B.19: *Skewness vs. forecast horizon for the case of Ireland and Denmark using the mean meteorological approach and the mean cluster approach*

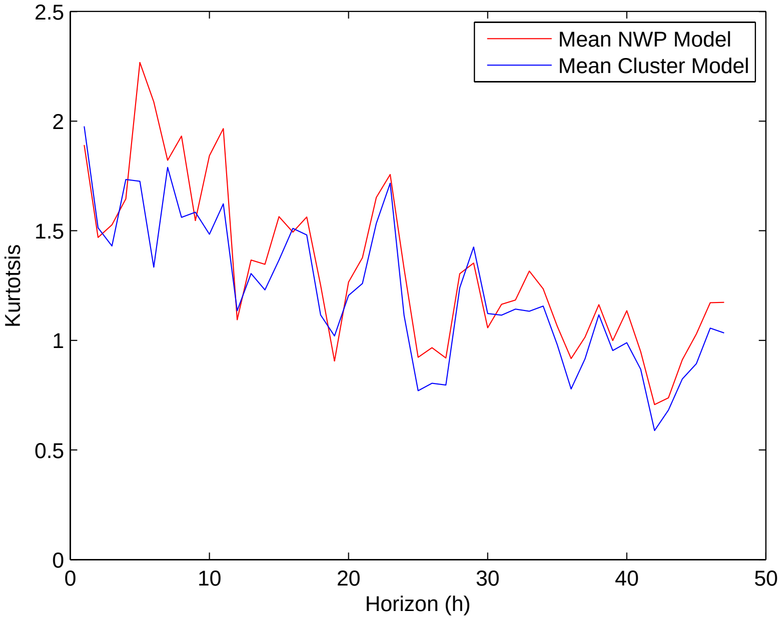


(a) Ireland

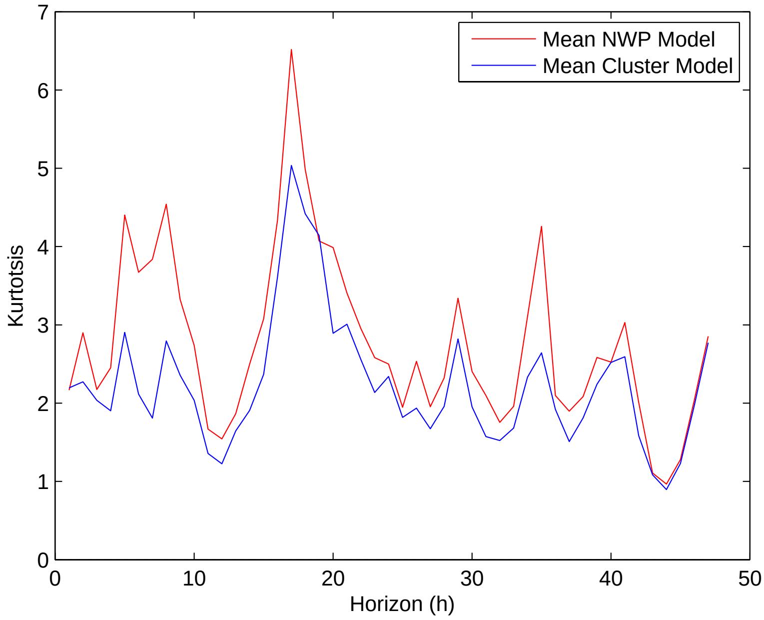


(b) Denmark

FIGURE B.20: *NSDE vs. forecast horizon for the case of Ireland and Denmark using the mean meteorological approach and the mean cluster approach*

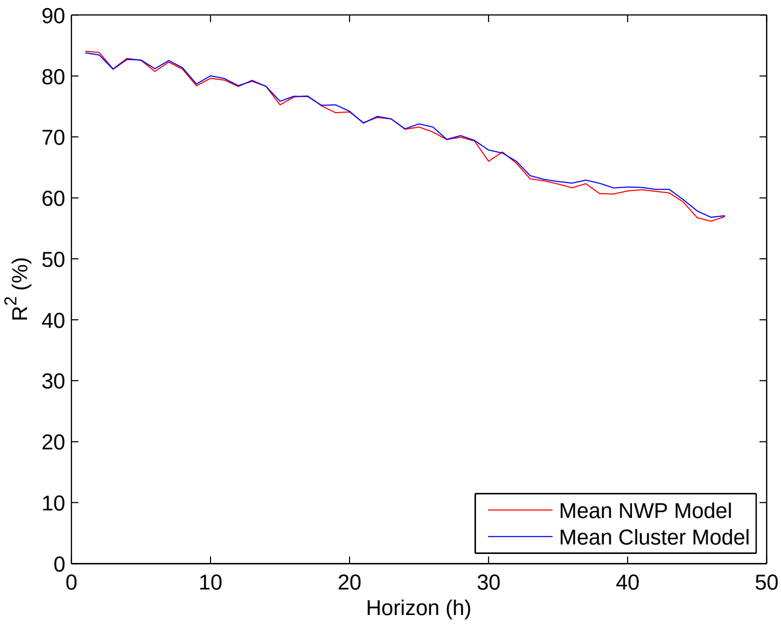


(a) Ireland

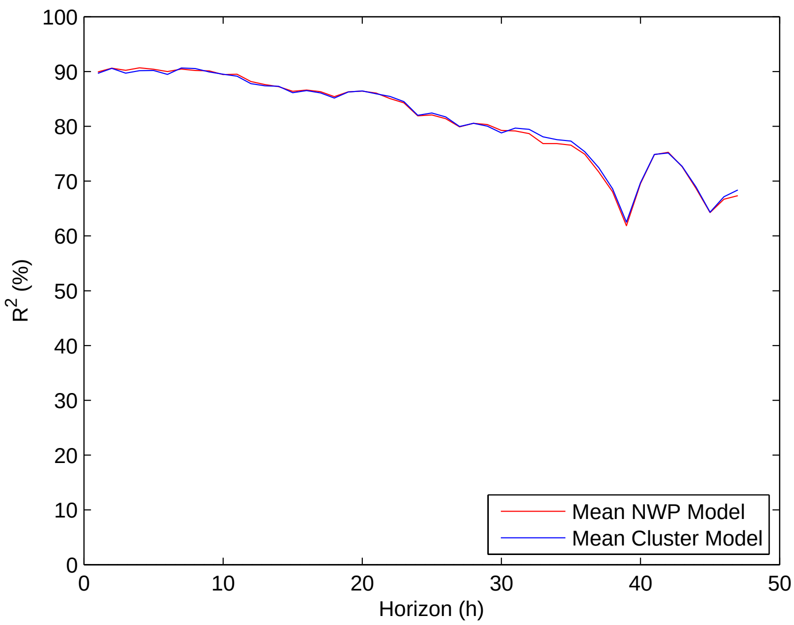


(b) Denmark

FIGURE B.21: *Kurtosis vs. forecast horizon for the case of Ireland and Denmark using the mean meteorological approach and the mean cluster approach*

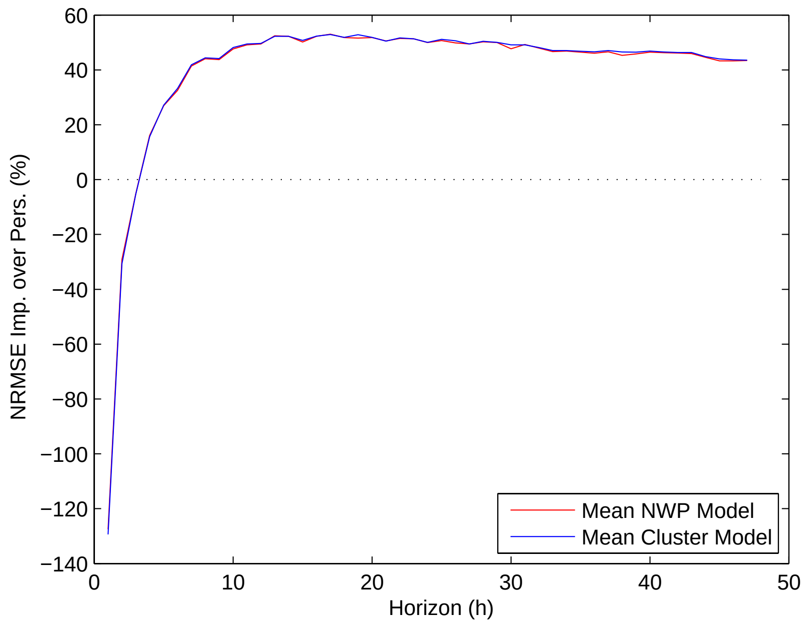


(a) Ireland

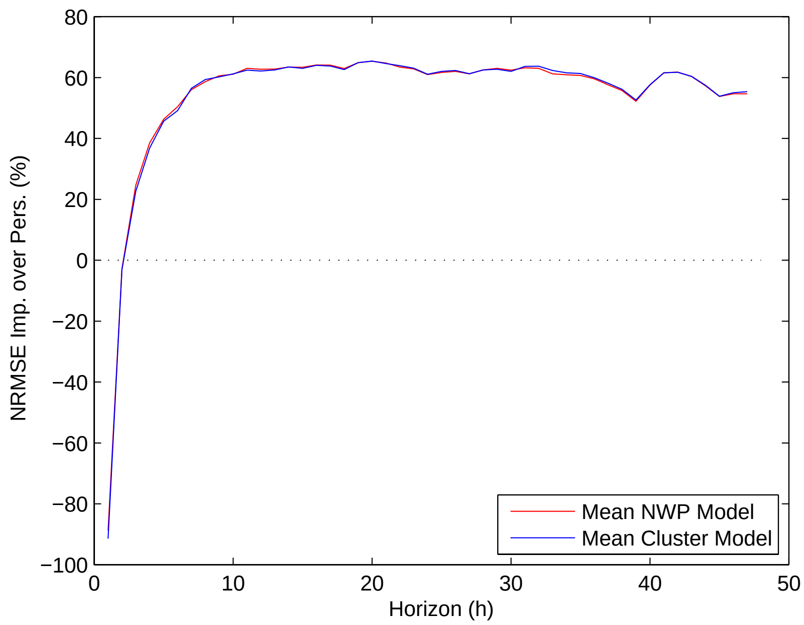


(b) Denmark

FIGURE B.22: *R^2 vs. forecast horizon for the case of Ireland and Denmark using the mean meteorological approach and the mean cluster approach*



(a) Ireland



(b) Denmark

FIGURE B.23: *NRMSE improvement over Persistence vs. forecast horizon for the case of Ireland and Denmark using the mean meteorological approach and the mean cluster approach*

APPENDIX C

Complementary Variable Combination Results

In the following the wind farm subsets having the lowest $\mathrm{NRMSE_{Sub}}$ for each subset cardinality are presented for the cases of Ireland and Denmark, for both the Persistence and NWP based forecast combinations. The geographical locations of the wind farms included in the subsets of cardinality 1 to 6 for the case of Denmark, for both the Persistence and NWP based forecast combinations are also presented. These results correspond to the study presented in section 5.4. The subsets are reported in the following tables:

- Subsets for the case of Ireland using Persistence forecast combinations (Table C.1).
- Subsets for the case of Ireland using NWP based forecast combinations (Table C.2).
- Subsets for the case of Denmark using Persistence forecast combinations (Table C.3).
- Subsets for the case of Denmark using NWP based forecast combinations (Table C.4).

The geographical locations of the wind farms are presented in the following figures:

- Wind farm locations for the Persistence combination subsets for the case of Denmark (Figure C.1).
- Wind farm locations for the NWP based combination subsets for the case of Denmark (Figure C.2).

TABLE C.1: *Subsets presenting the lowest* $\mathrm{NRMSE}_{\mathrm{Sub}}$ *for each cardinality, for the case of Ireland, considering Persistence wind farm forecasts.*

Cardinality	Wind farm indexes										
1	3										
2	3	8									
3	1	2	3								
4	1	2	3	6							
5	1	2	3	5	6						
6	1	2	3	5	6	8					
7	1	2	3	4	5	6	8				
8	1	2	3	4	5	6	8	10			
9	1	2	3	4	5	6	8	9	10		
10	1	2	3	4	5	6	7	8	9	10	
11	1	2	3	4	5	6	7	8	9	10	11

TABLE C.2: *Subsets presenting the lowest* $\mathrm{NRMSE}_{\mathrm{Sub}}$ *for each cardinality, for the case of Ireland, considering NWP based wind farm forecasts.*

Cardinality	Wind farm indexes										
1	2										
2	7	9									
3	4	5	9								
4	4	5	6	9							
5	4	5	6	9	10						
6	4	5	6	7	9	10					
7	1	3	4	5	6	7	9				
8	1	2	4	5	6	7	9	10			
9	1	2	3	4	5	6	7	9	10		
10	1	2	3	4	5	6	7	9	10	11	
11	1	2	3	4	5	6	7	8	9	10	11

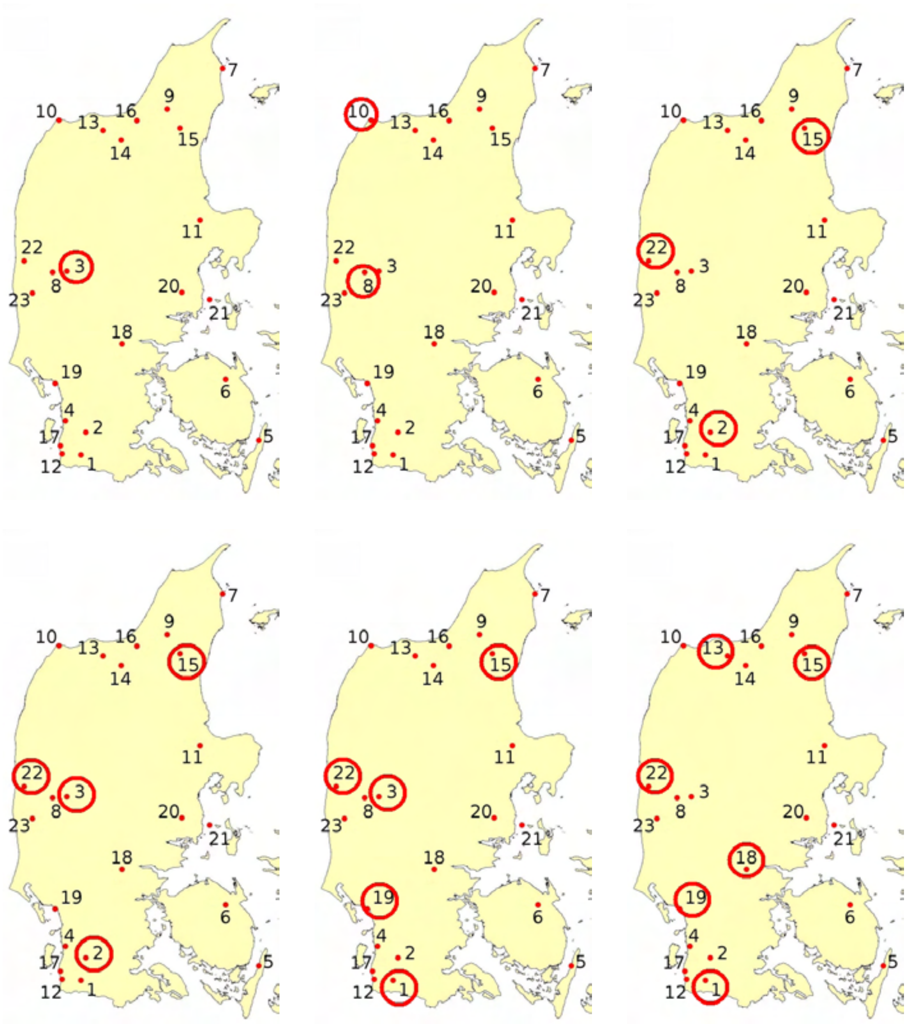


FIGURE C.1: *Locations of the wind farms whose production measurements belong to the best variable subsets, in terms of* $\mathrm{NRMSE}_{\mathrm{Sub}}$*, for cardinalities 1 to 6. The results are for the case of Denmark considering Persistence wind farm forecast combinations.*

TABLE C.3: *Subsets presenting the lowest* $\mathrm{NRMSE}_{\mathrm{Sub}}$ *for each cardinality, for the case of Denmark, considering wind farm Persistence forecasts.*

Cardinality	Wind farm indexes																						
1	3																						
2	8	10																					
3	2	15	22																				
4	2	3	15	22																			
5	1	3	15	19	22																		
6	1	13	15	18	19	22																	
7	1	11	13	15	18	19	22																
8	1	10	11	13	15	18	19	22															
9	1	9	10	11	13	15	18	19	22														
10	1	2	9	10	11	13	15	18	19	22													
11	2	5	6	9	10	11	13	15	18	19	22												
12	2	5	6	9	10	11	13	15	18	19	20	22											
13	2	5	6	9	10	11	13	15	18	19	20	22	23										
14	1	2	5	6	9	10	11	13	15	18	19	20	22	23									
15	1	2	5	6	9	10	11	13	15	16	18	19	20	22	23								
16	1	2	5	6	8	9	10	11	13	15	16	18	19	20	22	23							
17	1	2	5	6	8	9	10	11	13	14	15	16	18	19	20	22	23						
18	1	2	5	6	8	9	10	11	13	14	15	16	18	19	20	21	22	23					
19	1	2	4	5	6	8	9	10	11	13	14	15	16	18	19	20	21	22	23				
20	1	2	4	5	6	8	9	10	11	12	13	14	15	16	18	19	20	21	22	23			
21	1	2	3	4	5	6	8	9	10	11	12	13	14	15	16	18	19	20	21	22	23		
22	1	2	3	4	5	6	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
23	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23

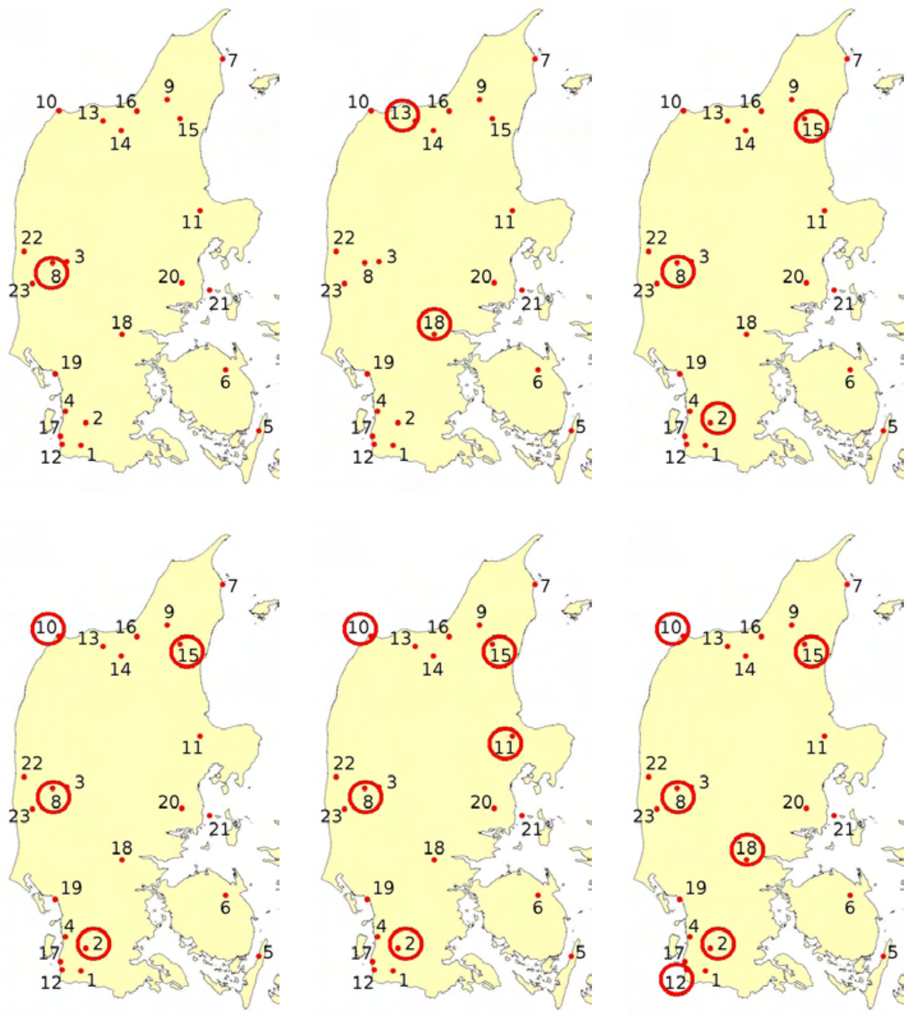


FIGURE C.2: *Locations of the wind farms whose NWP forecasts belong to the best variable subsets, in terms of* $\mathrm{NRMSE}_{\mathrm{Sub}}$*, for cardinalities 1 to 6. The results are for the case of Denmark considering NWP based wind farm forecast combinations.*

TABLE C.4: *Subsets presenting the lowest* $\mathrm{NRMSE}_{\mathrm{Sub}}$ *for each cardinality, for the case of Denmark, considering NWP based wind farm production forecast.*

Cardinality	Wind farm indexes																						
1	8																						
2	13	18																					
3	2	8	15																				
4	2	8	10	15																			
5	2	8	10	11	15																		
6	2	8	10	12	15	18																	
7	2	8	10	12	15	17	18																
8	2	8	10	12	15	17	18	22															
9	2	8	10	12	14	15	17	18	22														
10	2	7	8	10	12	14	16	17	18	22													
11	2	4	7	8	10	12	14	16	17	18	22												
12	2	3	4	7	8	10	12	14	16	17	18	22											
13	2	3	4	7	8	10	12	14	15	16	17	18	22										
14	2	3	4	7	8	10	12	14	15	16	17	18	19	22									
15	5	7	8	10	11	12	14	15	16	17	18	19	20	21	22								
16	2	3	4	7	8	10	11	12	14	15	16	17	18	20	21	22							
17	2	3	4	7	8	10	11	12	14	15	16	17	18	19	20	21	22						
18	2	3	4	7	8	10	11	12	13	14	15	16	17	18	19	20	21	22					
19	2	3	4	5	7	8	10	11	12	13	14	15	16	17	18	19	20	21	22				
20	2	3	4	5	7	8	10	11	12	13	14	15	16	17	18	19	20	21	22	23			
21	2	3	4	5	6	7	8	10	11	12	13	14	15	16	17	18	19	20	21	22	23		
22	1	2	3	4	5	6	7	8	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
23	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23

APPENDIX D

Traduction en Français

Cette annexe contient la traduction de l'Introduction, de la Conclusion et des chapeaux de chapitres.

D.1 Introduction

D.1.1 Contexte général

Des études récentes ont montré qu'au cours du siècle dernier, la température moyenne à la surface de la Terre a augmenté significativement. A ce titre, un rapport publié par le Groupe d'experts Intergouvernemental sur l'Evolution du Climat affirme que la température moyenne à la surface de la Terre a augmenté de $0.6 \pm 0.2°C$ au cours du vingtième siècle [1]. La même source rapporte que la concentration actuelle en dioxyde carbone et en méthane dans l'atmosphère est la plus élevée enregistrée en 420 000 ans. De plus, les concentrations de ces gazes ont augmenté respectivement de 31% et 151% depuis 1750. Dans cette étude, et d'autres, il est suggéré que l'augmentation des concentrations en gaz à effet de serre dans l'atmosphère a contribué au réchauffement global constaté. Qui plus est, comme l'augmentation de la concentration suit de près le développement de l'utilisation des énergies fossiles ces 150 dernières années, il a été suggéré que les émissions anthropogènes de gaz à effet de serre soient, en partie, responsables du réchauffement climatique. Ces découvertes ont poussée l'opinion publique à demander l'instauration de mesures visant à réduire les émissions de dioxyde de carbone dans plusieurs pays.

Une autre prise de conscience récente a été la non durabilité et les risques liés à l'utilisation d'énergies fossiles. Aujourd'hui, les énergies fossiles sont majoritaires dans le portefeuille énergétique mondial. Le pétrole est utilisé principalement dans le secteur du transport tandis que le gaz et le charbon sont principalement utilisés pour la génération d'électricité. Dans le futur, l'Agence Internationale de l'Energie prévoit une augmentation de 60% de la demande d'énergie d'ici à 2030 et un doublement de la demande d'électricité [2]. Bien que l'AIE ne prévoit pas de pique de production de pétrole dans les trois prochaines décennies, l'augmentation de la demande en énergies fossiles émanant des pays en développement, et dans une moindre mesure des pays développés, tendra les capacités de production. Les pays importateurs d'énergies fossiles deviendront de plus en plus dépendants d'un nombre décroissant de fournisseurs le plus souvent politiquement instables. Cette réduction du nombre fournisseurs exacerbera le risque de voir l'approvisionnement perturbé en des points névralgiques par lesquels sont acheminés le pétrole et le gaz.

Les inquiétudes grandissantes concernant le réchauffement climatique et la raréfaction des sources d'énergies conventionnelles ont poussé les instances dirigeantes à travers le monde à chercher des sources d'énergies plus propres (produisant moins de CO_2) et durables. Etant donné la croissance prévue de la demande mondiale d'électricité, une attention particulière a été apportée à la réduction de la consommation électrique, au développement de systèmes conventionnels de génération plus efficaces et à l'utilisation de sources d'énergies renouvelables.

Dans plusieurs pays des politiques visant à favoriser l'utilisation de moyens de conversion d'énergie efficaces et le développement de l'utilisation d'énergies renouvelables ont été mises en place [3]. Dans cette optique, l'Union Européenne a adopté en 2001 la Directive sur la promotion de l'électricité produite à partir de sources d'énergie renouvelables. Cette Directive fixe des objectifs d'utilisation d'énergies renouvelables pour la production d'électricité pour tous les pays membres. Plus précisément, cette Directive établie que 12% de la consommation d'énergie primaire et 22.1% de l'électricité produite dans l'Union en 2010 devra provenir de sources d'énergies renouvelables [4]. Au Etats-Unis d'Amérique, plus de quinze Etats ont mis en place des mesures visant à promouvoir le développement des sources d'énergie renouvelables [5]. Ces mesures sont le plus souvent des Renewable Portfolio Standards, des minima d'énergies renouvelables auxquels doivent se soumettre les gestionnaires de systèmes électriques. Un autre exemple est l'Inde où l'utilisation des principales énergies renouvelables (solaire, éolien, petit hydroélectrique et biomasse) est encouragée par différents mécanismes [6].

La plupart des sources d'énergies renouvelables peuvent être utilisées pour produire de l'électricité. Les technologies les plus couramment employées dans ce but sont : l'hydroélectricité, le solaire photovoltaïque, le solaire à concentration, l'éolien, la biomasse et la géothermie. Bien que la plupart de ces technologies aient atteint un degré de maturité certain, l'énergie éolienne est celle qui se développe la plus rapidement à l'heure actuelle. La capacité installée dans le monde a augmenté de 4800 MW en 1995 à plus de 74223 MW en 2006 [7]. Le Global Wind Energy Council (GWEC) prévoit qu'en 2010, 149500 MW seront installés dans le monde. Ce développement important de l'énergie éolienne peut s'expliquer par plusieurs facteurs.

Le premier de ces facteurs est la disponibilité de la ressource, il y a du vent partout sur la planète. Dans une étude réalisée par Hoogwijk [8], le potentiel éolien onshore économiquement réalisable, défini comme la quantité d'énergie éolienne pouvant être techniquement produite étant donné le coût de sources d'énergies alternatives, a été estimé au niveau mondial. Les résultats de cette étude montrent qu'une quantité équivalente à la consommation électrique mondiale de 2001 pourrait être produite pour moins de 0.07 US$/kWh. Cette étude compare aussi le rapport entre le potentiel techniquement exploitable et la consommation électrique de 1996 à un niveau régional. Des 17 régions définies dans l'étude, seules deux (l'Asie du Sud Est et le Japon) présentent un ratio inférieur à un. Un autre facteur jouant en faveur de l'éolien est son coût par rapport à celui d'autres énergies renouvelable. L'énergie éolienne est une des technologies renouvelables les plus compétitives [2]. Bien que d'autres technologies comme l'hydroélectricité et la géothermie soient plus compétitives, la disponibilité de celles-ci est limitée à certains territoires. Un troisième facteur vient s'ajouter aux deux précédents, le soutient politique important dont a bénéficié l'éolien. Des mesures de soutient efficaces ont été mises en place dans plusieurs pays. L'effet de celles-ci peut être constaté en analysant l'évolution du marché de l'éolien. En 2001, seuls 5 pays avaient atteint les 1000 MW d'éolien installés ; quatre ans plus tard, 11 pays avaient dépassé ce seuil [9].

Ce développement à grande échelle a conduit à une augmentation de la pénétration de l'éolien dans de nombreux réseaux électriques[1]. Dans certains réseaux, la pénétration éolienne en terme d'énergie peut atteindre des niveaux élevés, de 15% à 20 % [10]. Des niveaux aussi élevés peuvent se traduire par des niveaux de pénétration de 100% en termes de puissance pendant quelques heures par an, comme c'est le cas dans le réseau occidental du Danemark. D'aussi importants niveaux de pénétration peuvent poser des problèmes aux gestionnaires des réseaux car, contrairement aux sources conventionnelles d'énergie, l'éolien est variable et non contrôlable. Les gestionnaires de réseaux ne peuvent pas s'appuyer sur l'éolien pour maintenir l'équilibre du système. Pour intégrer de grandes quantités de puissance éolienne dans les réseaux électriques, la solution traditionnelle est d'augmenter les réserves de façon à pouvoir faire face aux variations imprévues de la puissance éolienne. Dans de nombreux réseaux, les réserves supplémentaires sont fournies par des centrales électriques utilisant des énergies fossiles. Par conséquent, ces réserves supplémentaires entraînent un surcoût d'opération du système et des émissions de polluants, réduisant ainsi les avantages escomptés de l'utilisation de l'énergie éolienne. Une solution possible à ce problème pourrait être l'utilisation de moyens de stockage d'énergie pour lisser les variations de l'éolien. Plusieurs études [11–14] se sont intéressées aux possibilités offertes par des moyens de stockage, ou d'autres sources d'énergie renouvelable comme l'hydroélectricité, pour atténuer les effets de la variabilité de l'éolien. Ces études montrent que de tels systèmes combinés peuvent être économiquement viables, en particulier dans le cadre du marché libéralisé de l'énergie. Toutefois, à notre connaissance, aucune centrale combinant éolien et stockage (ou autre source d'énergie renouvelable) n'a, à ce jour, été implantée à

[1] La pénétration éolienne est définie comme un ratio. Différents ratios ont été proposés dans la littérature : le rapport entre la puissance éolienne installée et la puissance installée totale d'un réseau, le rapport entre l'énergie éolienne et la demande totale d'énergie, ou encore le rapport entre la puissance éolienne injectée sur le réseau à un instant donné et la puissance totale injectée sur le réseau au même instant.

grande échelle. En effet, les coûts d'investissement des technologies de stockage pour des capacités comparables à celle de l'éolien (plusieurs MW) sont prohibitifs.

Pour maintenir l'équilibre du système, les gestionnaires de système électriques doivent planifier la disponibilité des moyens de production nécessaires pour répondre à la demande prévue. Des outils tels que des modèles de prévision de la demande, des algorithmes de unit commitment et d'economic dispatch sont utilisés pour optimiser les coûts de production. Du fait de sa nature variable et non contrôlable, l'éolien est traditionnellement considéré comme une demande négative [15–18]. Considéré ainsi, une possibilité pour atténuer certains des problèmes posés par l'éolien, est d'utiliser des prévisions de la puissance éolienne, de façon analogue aux prévisions de la demande, pour établir les plans de production et minimisé les coûts de production du système. De fait, les prévisions de la production éolienne sont nécessaires, voir obligatoires [19], dans des réseaux où la pénétration est élevée ou dans lesquels peu de moyens de réserves sont disponibles. Le besoin de prévisions de la production éolienne fiables est reconnu aujourd'hui par l'industrie des réseaux électriques comme un domaine crucial de recherche qui doit être soutenue pour permettre l'efficace intégration à grande échelle de l'éolien dans les réseaux électriques[20, 21].

D.1.2 Prédiction de la production éolienne

Connaître le futur a été un des désirs les plus chers de l'Homme depuis la nuit des temps. Ce désir a été une des forces motrices de la recherche scientifique dans de nombreux domaine. Depuis des siècles, des modèles astronomiques ont été développés pour prédire la position des astres. Dans d'autres disciplines, les lois de la physique ont été définies pour prédire le résultat de différents processus. Avec l'avènement de moyens abordables de calcul à très haute vitesse, les capacités de prédictions ont été étendues à des domaines tel que la météorologie où le besoin de calcul était une contrainte majeure [22]. De ce fait, il est aujourd'hui possible de prédire des processus complexes comme la production éolienne avec une précision raisonnable.

Le but de la prédiction de la production éolienne est de fournir une estimation de la puissance produite à un instant donné dans le futur. Plusieurs horizons de prédictions peuvent être considérés. Le cas auquel nous nous intéressons est la prédiction à court terme dans lequel des prévisions pour les prochaines 48 à 120 heures, avec une résolution horaire, doivent être fournies. Ces horizons et cette résolution sont considérés car ils sont traditionnellement utilisés dans le cadre des calculs de unit commitment [23] ou pour la participation aux bourses de l'électricité. Les prédictions les plus couramment fournies par des modèles de prédiction sont des prédictions " point " : pour chaque pas de temps dans le futur, une seule valeur de puissance est fournie. Toutefois, des modèles de prédiction probabilistes sont en cours de développement. Ceux-ci fournissent des informations supplémentaires sur la distribution probable de la puissance éolienne produite.

Pour calculer des prévisions à court terme de la production éolienne, il est nécessaire d'associer les sorties de modèles numériques de prévision météorologique et un modèle de conversion. Ce dernier doit être capable de déterminer la production probable d'un parc éolien à partir des prévisions météorologiques et de mesures de la production et/ou de variables météorologiques. Ces modèles de conversion sont couramment appelés modèles de prédiction de la production éolienne. Différentes approches de conversion existent. Toutefois, d'un point de vue général, ces modèles peuvent être considérés comme des fonctions qui modélisent la production des parcs éoliens en fonction de la situation météorologique attendue. De ce point de vue, une bonne connaissance des phénomènes gouvernant la production d'un ensemble de turbines est nécessaire pour établir des modèles qui puissent précisément prédire le comportement de ces dernières.

La plupart des modèles développés jusqu'à présent sont principalement destinés à la prédiction de la production de parcs éoliens, qui sont le plus souvent considérés comme des centrales individuelles. A mesure que le nombre de parcs éoliens augmente, les utilisateurs des prévisions de production expriment le besoin de disposer de prévision de la production agrégée de tous les parcs situés dans une zone géographique définie. Dans le cadre de la gestion des systèmes électriques, des prévisions de la production agrégée sont nécessaires pour plusieurs fonctions : la gestion des congestions réseau, l'estimation des réserves nécessaires, la planification des échanges avec d'autres réseaux, etc. La prédiction de la production éolienne régionale peut donc être définie comme la prédiction de la production agrégée de tous les parcs éoliens dispersés dans une zone géographique définie.

D.1.3 Le prédiction de la production éolienne régionale

La prédiction de la production éolienne régionale s'intéresse à la production agrégée des parcs éoliens situés dans zone géographique définie. La prédiction de la production régionale se distingue de la prédiction de la production d'un parc seul par plusieurs aspects. Le premier est que la production régionale n'a pas les mêmes caractéristiques que la production d'un parc. Un deuxième aspect est que l'information disponible dans le cadre de la prévision régionale n'est pas nécessairement aussi complète que celle disponible pour des modèles de prévision pour des parcs individuels. Du fait de ces différences, des modèles spécialisés sont nécessaires pour fournir les meilleures prédictions possibles de la production éolienne régionale.

La principale caractéristique de la production éolienne régionale est que la capacité installée considérée est répandue dans une zone géographique étendue. Cette dispersion a pour principal avantage de lisser la production agrégée des parcs éoliens. En sommant la production de plusieurs parcs éoliens, les variations individuelles de puissances se compensent et la production agrégée résultante présente des variations plus lentes. Cet effet de lissage statistique nous mène à la principale caractéristique de la production éolienne régionale qui est que la production géographiquement dispersée peut être prédite plus précisément que la production individuelle des parcs éoliens [24].

Pour prédire la production éolienne au niveau régionale ou nationale, l'idéal serait de prédire la production de chaque parc individuellement pour obtenir par sommation la production totale pour les heures à venir. Cependant, les modèles de prédiction éolienne requièrent des données concernant les parcs éoliens pour réaliser leurs calculs. Ces données sont de deux types : statiques et dynamiques. Les données statiques sont des données qui décrivent le parc éolien : la puissance nominale, les courbes de puissance des turbines, etc.

Ces informations sont généralement considérées comme n'évoluant pas dans le temps. Les données dynamiques, telles que les mesures de production et les prévisions de variables météorologiques, sont des séries temporelles qui évoluent dans le temps.

Dans des réseaux électriques comportant un nombre important de parcs éoliens, prédire la production de chaque parc individuellement peut s'avérer impossible en raison d'un manque de données adéquates sur les parcs (données statiques) et/ou de mesures de production (données dynamiques). Par exemple, des mesures de production ne sont pas toujours disponibles pour tous les parcs. Ceci s'explique par le fait que des systèmes de télémesure (Supervisory Control and Data Aquisition, SCADA) ne sont pas installés dans tous les parcs. Il est à noter que dans plusieurs pays les règles de connexion imposent l'installation de tels équipements pour les parcs dont la puissance nominale dépasse un certain seuil.

L'absence de systèmes SCADA dans certains parcs implique que la production éolienne totale, en termes de puissance, est rarement connue précisément. Toutefois, même si la puissance que chaque parc injecte sur le réseau peut être connue avec un certain délai, la mesure de la capacité disponible à un instant est difficile à obtenir. Ici la capacité disponible est la somme des capacités installées des turbines en fonctionnement dans la région à un moment donné. Ce manque d'information peut être particulièrement handicapant pour l'apprentissage de modèles de prédiction statistiques. En effet, la mesure de la production ne peut pas être corrigée de la capacité disponible ce qui entraîne la présence de bruit dans les données servant à l'apprentissage. Ce problème est similaire à celui rencontré lorsque des prévisions de maintenance ne sont pas disponibles lors des calculs de prédiction de production éolienne.

En raison de ce manque habituel de données statiques et particulièrement de données dynamiques, l'approche la plus courante de la prédiction de la production éolienne régionale est l'utilisation de méthodes d'extrapolation, dites d'upscaling. Ces méthodes utilisent les données statiques et dynamiques disponibles pour extrapoler la meilleure prédiction possible de la production des parcs se trouvant dans une zone géographique. Dans le reste de ce document, les parcs éoliens pour lesquels des données dynamiques sont disponibles seront appelés parcs de référence.

D.1.4 But de la thèse

Le but du travail présenté dans cette thèse est de développer un cadre et de fournir des conseils qui pourront être utiles à des prévisionnistes souhaitant implanter des modèles de prédiction de la production éolienne régionale. Plus précisément, le but est de fournir aux concepteurs de modèles statistique des réponses aux questions suivantes : quelles stratégies de modélisation existent et quelles sont leurs influences sur la précision des prédictions ? Dans quelle mesure le choix du modèle statistique utilisé influence la précision des prédictions ? A quel point le choix des variables explicatives influence la performance des modèles et comment effectuer la sélection de ces variables explicatives ?

Pour fournir des réponses à ces questions, nous commencerons par proposer un cadre pour la caractérisation de la production éolienne régionale en tant que série temporelle.

De ce point de vue, la production éolienne régionale peut être abordée comme une série temporelle simple ou comme un agrégat de séries temporelles représentant les productions de parcs éoliens individuels. Ensuite, nous étudierons la relation qui existe entre les prévisions météorologiques et la production éolienne régionale. Ainsi, des aspects saillant du problème de prédiction de la production éolienne régionale, devant être pris en compte lors de la modélisation, seront identifiés. Cette caractérisation fournira aussi des résultats qui permettront une étude plus détaillée des performances des modèles de prédiction de la production éolienne régionale.

Nous examinerons ensuite le problème de prédiction de la production régionale du point de vue de l'apprentissage statistique. Après une revue des solutions proposées dans la littérature sur la prédiction de la production éolienne régionale, nous définissons trois approches de bases qui peuvent être utilisées pour combiner des modèles (le plus souvent des modèles de prédiction de production de parcs individuels) pour construire des modèles régionaux. Nous nous concentrons ensuite sur l'examen de l'impacte de ces différentes approches sur la précision des prédictions. Le but est de déterminer si une de ces approches est supérieure aux autres. A partir de là, nous examinerons l'influence du choix du modèle statistique de base sur les performances du modèle de prédiction régionale obtenu par combinaison de modèles de base. Pour ce faire, nous comparerons les performances obtenues en utilisant deux modèles de base, utilisés dans différentes approches de combinaison. Cette comparaison est rendue possible par l'introduction d'un nouveau modèle statistique. Nous démontrons ainsi que les performances de ce nouveau modèle sont comparables à celles de modèles de l'état de l'art. A partir de ces deux examens nous concluons sur l'influence relative de l'approche de combinaison et du choix du modèle statistique de base, sur la précision des prédictions de la production éolienne régionale.

Bien que le choix de modèle statistique et le choix de l'approche de combinaison aient un impact sur la performance des modèles de prédiction de la production éolienne régionale, une part important de la précision, ou de l'imprécision, des modèles est due aux variables explicatives utilisées pour modéliser le processus. Pour examiner l'impact des variables explicatives sur la performance des modèles prévision, nous commençons par caractériser les deux types de variables utilisées dans la prédiction statistique de la production éolienne : les mesures de production et les prévisions de variables météorologiques. L'influence de ces deux types de variables est ensuite étudiée dans le cadre de la prédiction de la production éolienne régionale, où le nombre de variables potentiellement disponible expose les modèles statistiques à la malédiction de la dimension. Plus précisément, la précision des prédictions obtenues avec toutes les combinaisons de variables explicatives disponibles est analysée. De cette étude, des règles de conduite générales concernant le nombre de variables explicatives à considérer en fonction de leur type sont établies. De plus, étant donné le nombre important de variables pertinentes et souvent redondantes disponibles dans le cadre de la prédiction de la production éolienne régionale, des techniques de sélection de variable explicatives sont étudiées. Deux techniques de sélection de variables, sont proposées. Ces techniques reposent sur des caractéristiques propres au problème de la prévision de la production régionale. Ces deux méthodes originales sont comparées à une méthode de l'état de l'art et les performances des trois sont étudiées de façon détaillée. Ces trois méthodes constituent une première réponse au problème de la sélection de variables explicatives dans le cadre de la prédiction de la production éolienne régionale.

D.1.5 Structure de la thèse

Pour introduire le lecteur au problème de la prédiction de la production éolienne le Chapitre 2 présente l'état de l'art de la prédiction de la production éolienne. Dans cette revue de la littérature, les fondements de la prédiction de la production éolienne sont abordés. Un exposé général des modèles de prédiction existants est présenté ainsi que les critères communément utilisés pour l'évaluation de leur performance. Les développements récents dans les domaines de l'estimation de l'incertitude de la prédiction et de la valeur économique des prédictions de la production éolienne sont rapportés. Notre revue se concentre ensuite sur les modèles de prédiction de la production éolienne régionale proposés dans la littérature. De plus, les lacunes dans le domaine de la prédiction de la production éolienne régionale sont identifiées et les objectifs de la thèse sont justifiés.

Le but du Chapitre 3 est de caractériser la production éolienne régionale et sa relation avec les variables explicatives disponibles. Cette caractérisation servira de base aux développements ultérieurs de la thèse. Le chapitre débute par la formalisation du problème de la prédiction de la production éolienne régionale. Un cadre théorique est ensuite proposé pour étudier la production éolienne régionale et ses liens statistiques avec les variables explicatives disponibles ; dans le cas présent des prévisions issues de modèles météorologiques. Ce cadre considère la production éolienne régionale comme une série temporelle simple et comme un agrégat de séries temporelles, permettant ainsi l'étude de la série de deux points de vue. La relation statistique entre la production régionale et les prévisions météorologiques est caractérisée en utilisant une mesure de l'information mutuelle existant entre les différentes variables. Ce cadre de caractérisation est appliqué à deux cas d'étude : un cas comprenant 11 fermes en Irlande et un autre cas comprenant la production éolienne totale de la péninsule du Jütland et de l'île de Funen, au Danemark. Les résultats de la caractérisation des deux sont analysés et comparés. De cette analyse, des conclusions générales sur les propriétés les plus saillantes du problème de la prédiction de la production éolienne régionale sont présentées.

Dans le Chapitre 4 le problème de la prédiction de la production régionale est abordé du point de vue de l'apprentissage statistique. De ce point de vue, les modèles existants sont examinés et trois méthodes de combinaison de modèles de base (ou sous-modèles) sont identifiées. Nous nous attachons ensuite à caractériser la performance obtenue avec différents sous-modèles dans ces trois méthodes. Pour ce faire, deux modèles de base sont utilisés : un modèle de réseau de neurones floues (F-NN) de l'état de l'art développé par Kariniotakis [25] et un modèle développé dans le cadre de cette thèse, le modèle Regressive Power Curve (RPC). En premier lieu, les performances du modèle RPC sont validées en les comparant à celles d'autres modèles de l'état de l'art, sur trois cas d'études représentatifs de terrains de complexités différentes. Nous montrons que les performances du modèle RPC sont comparables à celles des autres modèles de l'état de l'art. Les deux sous-modèles

(F-NN et RPC) sont ensuite utilisés pour étudier les caractéristiques des trois méthodes de combinaison pour la prédiction de la production éolienne régionale et pour étudier l'influence de l'usage de différents sous-modèles dans celles-ci. L'évaluation est conduite sur les cas d'étude Irlandais et Danois décrits précédemment.

Le problème général de l'impact des variables explicatives sur la précision des prédictions est abordé dans le Chapitre 5. Dans la première partie, nous proposons une méthodologie pour étudier l'influence de différents types de variables explicatives sur la précision des prédictions de la production éolienne régionale. Cette méthodologie est ensuite appliquée à des variables explicatives telles que les séries temporelles de mesures de production et aux prévisions de la vitesse du vent. De l'analyse détaillée des résultats, des conclusions générales sont tirées et la nécessité de mener une sélection des variables explicatives est démontrée. Ceci nous mène à la deuxième partie du chapitre qui s'intéresse à l'utilisation de méthodes automatiques de sélection de variables explicatives dans le cadre de la prédiction de la production éolienne régionale. Cette partie commence par un exposé général des différentes méthodologies de sélection existantes. De cette revue, trois méthodes de sélection automatique sont proposées. La première est une méthode de l'état de l'art que nous utilisons comme méthode de référence. Les deux autres méthodes, une méthode filtre basée sur l'agrégation de variables et une méthode wrapper basée sur la stabilité de transition, sont proposées. Ces deux dernières sont des adaptations de méthodologies existantes et exploitent certaines caractéristiques de la prédiction de la production éolienne régionale misent en évidence dans la première partie du chapitre. Les trois méthodes sont comparées sur les cas d'étude Danois et Irlandais. De cette comparaison, une analyse détaillée des propriétés des méthodes est réalisée et des conclusions générales sur leur utilité sont obtenues.

Finalement, le Chapitre 6 fourni des conclusions générales et des conseils concernant les sujets abordés dans cet ouvrage. De ces conclusions, de futures voies de recherche sont identifiées et discutées.

D.2 Conclusions générales

D.2.1 Conclusions et contribution de la thèse

Lorsque les travaux présentés dans cette thèse ont débutés il y a quelques années, seuls quelques réseaux électriques se trouvaient confrontés à une importante pénétration éolienne. A cette époque, peu de modèles de prévision régionaux étaient décris dans la littérature, la plupart étant des solutions spécifiques mises en place dans quelques systèmes électriques. La plupart des travaux sur la prévision de la production éolienne régionale se concentraient sur la mise en place de modèle et sur l'examen des bénéfices que ceux-ci tirent du lissage statistique du au foisonnement éolien, qui conduit à une réduction des erreurs de prédiction. Aucune règle de mise en œuvre n'était alors disponible pour guider les prévisionnistes souhaitant mettre en place des modèles de prédiction régionaux. Dans les travaux présentés dans cet ouvrage, la prédiction de la production éolienne régionale a été étudiée d'un point de vue général. Nous avons montré que différents cas de prédiction régionale partagent des propriétés communes et que ces propriétés peuvent être exploitées lorsque l'on implante des solutions de prédiction de la production éolienne régionale. En ce qui concerne les aspects de modélisation, nous avons montré que peu de différences existent entre diverses approches de modélisation. De plus, nous avons montré que l'approche consistant à combiner de nombreuses prévisions de production de parcs individuels, ou à exploiter un nombre grandissant de variables explicatives ne conduit pas à une amélioration sans limite de la précision des prédictions. En fait, nous avons montré que seul un petit nombre de variables explicatives, bien choisi, est nécessaire pour prédire la production éolienne régionale de façon adéquate.

De façon à aborder le problème de la prédiction éolienne régionale d'un point de vue général, il était nécessaire de le caractériser. Nous avons les caractéristiques de la production éolienne régionale et nous avons montré que la production régionale est un processus non stationnaire. La non-stationnarité du processus peut être contrecarrée par l'utilisation de modèles de prévisions adaptatifs qui re-estiment leurs paramètres périodiquement. Nous avons aussi montré que la production éolienne régionale présente une forte auto-corrélation. Pour de tels processus et pour des horizons à très court terme, des modèles autorégressifs, dont la Persistance, présenteront des performances acceptables. Cette auto-corrélation élevée provient en partie du lissage statistique dont bénéficie la production régionale en raison de l'agrégation de productions de parcs éoliens individuels. Pour caractériser cet effet de lissage, nous avons examiné théoriquement et expérimentalement les facteurs qui le gouvernent. De cette caractérisation, nous concluons que le lissage statistique est proportionnel à la taille de la région considérée et au degré de dispersion des parcs éoliens qui s'y trouvent. Ces conclusions corroborent les résultats expérimentaux décris dans [24].

De plus, nous avons montré que le nombre de fermes situées dans une région n'influence pas directement le lissage observé. Toutefois, quand le nombre de fermes installées dans une région augmente, la probabilité de constater du lissage augmente. Ayant conduit une étude approfondie des propriétés de la production éolienne régionale, nous nous sommes intéressés aux relations existant entre cette dernière et les variables explicatives que sont les productions des parcs éoliens individuels et les prévisions météorologiques. Nous avons montré qu'il est possible d'estimer la production régionale à partir d'un petit de nombre de mesures de production de parcs individuels et que la relation entre la production régionale et la production des parcs éoliens qui la composent est hautement linéaire. Nous avons aussi montré que la relation entre les productions des parcs éoliens et la production régionale est un processus non-stationnaire, ce qui plaide en faveur de l'utilisation de modèles de prévision auto-adaptatifs capables de faire de la sélection de variables explicatives en cours d'exécution. En ce qui concerne les prévisions de variables météorologiques, nous avons montré que, logiquement, ce sont les prévisions de vitesse de vent qui fournissent l'information la plus utile concernant la production régionale future. Nous avons aussi mis en évidence la décroissance du pouvoir explicatif des variables météorologiques à mesure que l'horizon de prédiction augmente. Finalement, nous avons montré que les points de la grille du modèle météorologique ne fournissent pas une quantité égale d'information sur la production future ; certains fournissent significativement plus d'information que d'autres.

Après avoir caractérisé la production régionale, nous avons continué notre étude en nous intéressant aux propriétés des modèles de prévision statistiques dans le cadre de la prédiction éolienne régionale. Nous avons montré que des sous-modèles de prévision peuvent être combinés de trois manières basiques, à partir desquelles une infinité de configurations de modèles peuvent être obtenues. Ayant définie ces trois approches de combinaison, nous avons étudié à quel point celles-ci mènent à des performances de prédiction différentes et si l'utilisation de sous-modèles différents dans ces approches conduit à des différences de performance.

Afin de répondre à ces questions, nous avons défini un nouveau modèle de prévision, RPC, que nous avons comparé à d'autres modèles sur plusieurs cas d'étude de la littérature. Nous avons montré que ce modèle à des performances comparables à ceux de l'état de l'art, et un avantage : la définition de son architecture et l'estimation de ses paramètres sont automatiques et très rapides. Etant donné sa simplicité, ce modèle pourrait être utilisé comme modèle de référence dans de futures études. Comparer les performances de nouveaux modèles à celles de RPC permet une évaluation plus significative des avancées que leur comparaison à des modèles naïfs comme la Persistance. En utilisant le modèle RPC et le modèle F-NN, nous avons étudié la qualité des prévisions obtenues avec les trois approches de combinaison de sous-modèles. De cette étude, nous pouvons conclure qu'en termes de performance peu de différences existent entre les approches. La sélection de l'approche de combinaison devrait, à notre avis, répondre à un besoin pragmatique : fournir les prévisions utiles à l'utilisateur. En fonction des besoins de l'utilisateur et des données dont il dispose (prévision météorologiques, mesure de production, prévision de production de parcs éoliens) une combinaison raisonnable peut être mise en œuvre pour fournir des prédictions utiles et précises. Lors de la conception d'un tel outil, le prévisionniste devrait : garder à l'esprit les besoins de l'utilisateur, les contraintes spécifiques du problème de prédiction éolienne régionale auxquelles il est confronté et limiter, autant que possible, la complexité et la dimensionnalité de la combinaison de sous-modèles. Ces guides de conception se trouvent renforcés par les résultats de la comparaison des performances des modèles RPC et F-NN : les deux modèles présentent des performances très similaires et aucun ne surpasse systématiquement l'autre. Il n'existe probablement pas de sous-modèle qui soit le plus adapté à la prévision régionale. C'est plutôt l'habileté du prévisionniste et sa connaissance des propriétés du sous-modèle qui mèneront à des différences en termes de performance. Le prévisionniste devrait choisir un modèle dont le pouvoir de modélisation est suffisamment élevé et dont il maîtrise les processus de définition d'architecture et d'estimation des paramètres. Ainsi, le prévisionniste pourra se concentrer sur l'examen des particularités du problème auxquelles il est confronté.

Pour fournir aux prévisionnistes plus d'informations sur la conception de modèles de prévision de la production éolienne régionale, nous avons abordé le problème de la sélection des variables explicatives fournissant au modèle l'information la plus utile concernant le processus en question. Afin d'atteindre cet objectif, nous avons proposé un cadre d'examen de l'impact de l'utilisation de différentes variables explicatives sur la précision des prédictions obtenues. Nous avons montré que les mesures de production fournissent de l'information essentielle sur l'évolution à court terme de la production régionale, tandis que les prévisions météorologiques, en particulier les prévisions de la vitesse du vent, sont primordiales pour les prévisions à moyen et long terme. Nous avons aussi montré que la quantité de bruit présent dans les mesures de production et dans les prévisions de vitesse de vent peut limiter le nombre de variable de chaque type qu'un modèle peut efficacement exploiter. Etant donné que les mesures de production sont peu bruitées, les modèles de prévision peuvent en exploiter un plus grand nombre. Ainsi, il est possible de fournir à un modèle une estimation précise de la situation météorologique et du niveau de production régionale au moment où les prédictions de production sont calculées. A l'inverse, les prévisions de vitesse de vent sont beaucoup plus bruitées et, étant donné qu'elles proviennent le plus souvent du même modèle météorologique, elles présentent des interdépendances statistiques plus élevées. Le niveau de bruit plus élevé implique qu'un modèle de prévision de production peut exploiter un plus petit nombre de ces variables. Aussi, les dépendances entre ces variables signifient qu'elles sont, dans une certaine mesure, redondantes. Cette redondance limite aussi le nombre de ces variables qu'un modèle peut exploiter avec profit. En effet, le fait d'augmenter le nombre de variables météorologiques prédites fournies au modèle de prédiction de la production n'entraîne pas une augmentation proportionnelle de l'information sur la situation météorologique future fournie à ce dernier. Nos résultats montrent que relativement peu de prévisions de vitesses de vent (ente 5 et 8) sont suffisantes pour prédire efficacement la production éolienne régionale. L'utilisation d'un plus grand nombre de variables ne se traduit pas par une augmentation de la précision de prédiction de production et peut même se traduire par une dégradation de la précision.

Ces résultats ont mis en lumière le besoin de procéder à la sélection des variables explicatives utilisées par un modèle de prévision de la production éolienne. Il est non seulement nécessaire de limiter le nombre de variables utilisées, mais dans le cadre de la prévision de la production éolienne régionale, le prévisionniste est confronté à un nombre de variable d'un ordre de grandeur supérieur au nombre disponible dans le cas de la prévision de la production d'un parc éolien. Afin d'aider le prévisionniste, nous avons examiné les possibilités offertes par des méthodes automatiques de sélection de variables dans le cadre de la prédiction de la production éolienne régionale où toutes les variables disponibles sont pertinentes vis à vis du processus modélisé. A partir des propriétés de la production régionale mises en évidence par la caractérisation et des résultats sur l'influence des variables sur la précision des prédictions de production, nous avons proposé deux heuristiques de sélection : une méthode basée sur l'agrégation de variables et une basée sur la méthode wrapper. Cette dernière exploite la stabilité de transition observée entre les meilleurs sous-ensembles de variables de cardinalités différentes. Ces deux heuristiques ont été comparées à une heuristique de l'état de l'art, l'algorithme MIFS qui est basé sur l'utilisation d'une mesure de l'information mutuelle entre variables. Nous avons montré que ces trois heuristiques permettent la sélection de sous-ensembles de variables explicatives qui mènent à des prédictions de production ayant de bons niveaux de précision. De plus, ces trois heuristiques présentent des coûts de calcul différents. L'heuristique basée sur l'agrégation de variable est la moins coûteuse en ce sens, mais ne permet pas une sélection aussi bonne que les deux autres en termes de précision des prédictions obtenues. La méthode MIFS permet une bonne sélection de variable avec un temps de calcul raisonnable et un coût de calcul correct. Finalement, la méthode wrapper permet la meilleure sélection mais requière le plus de temps de calcul. D'après les résultats obtenus avec ces trois heuristiques, il apparaît clairement que la sélection de variables peut être avantageusement automatisée dans le cadre prédiction de la production éolienne régionale. De plus, les deux nouvelles méthodes proposées et la méthode MIFS constituent une première réponse au problème de la sélection de variables explicatives pour le problème de la prédiction de la production éolienne régionale.

D.2.2 Perspectives

Le but des travaux présentés dans cet ouvrage était de fournir aux prévisionnistes confrontés au problème de la prédiction de la production éolienne régionale des guides concernant les problèmes spécifiques qui doivent être traités dans ce cadre. Nous avons abordé différentes approches de modélisation ainsi que l'utilisation de différents modèles statistiques. Nous avons aussi proposé des méthodes automatiques de sélection des variables explicatives qui permettent d'éviter la malédiction de la dimension. Toutefois, d'autres aspects du problème de la prédiction de la production éolienne régionale n'ont pas été abordés.

Un de ces aspects est la possibilité de rendre les modèles de prédiction de la production régionale auto-adaptatifs afin de leur permettre de mieux gérer la non-stationnarité de la production éolienne régionale. Ceci peut paraître une étape simple à réaliser étant donné le grand nombre de modèles statistiques auto-adaptatifs proposés dans la littérature ; les modèles F-NN et RPC utilisés dans nos travaux peuvent être rendu auto-adaptatifs sans difficulté. Le cœur du problème est que pour rendre un modèle adaptatif, il est nécessaire de pouvoir calculer ses erreurs de prédiction. Ceci implique que la production régionale soit mesurée et que cette mesure puisse être fournie aux modèles. Cependant la production régionale est rarement mesurée en temps réel, ou en léger différé. Dans des pays où la puissance éolienne installée est élevée, des mesures provenant d'un grand nombre de parcs éoliens doivent être récupérées et la production horaire est, le plus souvent, connue avec un retard de plusieurs semaines. De ce fait, l'adaptation des modèles sera nécessairement retardée. Cet état de fait soulève plusieurs questions : dans quelle mesure ce retard affecte la précision de prédictions ? Quel est le retard maximum au-delà duquel les prévisions ne sont pas améliorées et à quel point les prévisions peuvent elles être améliorées en réduisant ce retard ? Des réponses à ces questions permettraient de quantifier le compromis entre le retard de mesure et la qualité des prédictions. Ceci permettrait par conséquent aux décideurs d'effectuer un arbitrage entre le coût d'une réduction du retard et le bénéfice économique dérivé de prédictions plus précises.

Une autre facette du problème de prédiction de la production éolienne régionale qu'il conviendra d'étudier est la possibilité de produire des prévisions probabilistes de la production régionale. Au cours des dernières années, un effort de recherche important a été consacré à la création de modèles de prévision probabilistes pour la production des parcs éoliens. Fournir le même type de prévision pour la production régionale pourrait être utile aux gestionnaires de réseaux électriques souhaitant exploiter au maximum les prévisions de production dans leur processus de gestion, i.e. unit commitment, economic dispatch, etc. Dans ce cadre, la caractérisation des erreurs de prédiction pour un seul parc présentée dans [36] pourrait être étendue aux erreurs de prédiction de la production régionale. A partir d'une telle caractérisation, des modèles probabilistes régionaux basés sur des modèles tel que celui décrit dans [145] pourrait être développés. Etant donné le lissage statistique des erreurs qui a lieu dans le cadre régional, il est probable que les distributions de production obtenues soient proportionnellement plus pointues que celles observées dans le cadre de la prédiction des productions de parcs individuels. Ceci pourrait à son tour mener au développement d'algorithmes stochastiques de gestion de réseaux électriques, comparable dans leurs principes à celui proposé dans [147]. De tels algorithmes pourraient être utilisés pour minimiser le surcoût subi par des réseaux fortement pénétrés par l'éolien.

Dans la même direction, les améliorations des prédictions régionales qui pourraient être obtenues par l'utilisation d'ensembles météorologiques devraient être évaluées. L'utilisation de ce type de prévisions météorologiques a déjà été étudiée pour la génération de prévisions probabilistes améliorées de la production de parcs individuels. La même possibilité existe pour la production éolienne régionale ; en particulier puisqu'il a été démontré que la moyenne des membres d'un ensemble conduit à des prévisions de la production de parc plus précises que celles obtenues en utilisant le membre de control [148]. Lors de l'utilisation d'ensembles dans le cadre régionale, il sera toutefois nécessaire de veiller à extraire l'information supplémentaire contenue dans les ensembles sans pour autant exposer le modèle de prévision à des entrées de dimension trop élevées.

D'un point de vue plus général, le cadre d'analyse décrit dans cet ouvrage devrait être appliqué à un plus grand nombre de cas d'études, portant sur des régions plus grandes et plus variées. Des régions telles que l'Espagne ou la France sont dignes d'intérêt étant donné la variété des différentes situations météorologiques et la diversité des complexités de terrains qu'elles présentent. Avec de tels cas l'effet bénéfique résultant d'une plus faible corrélation entre les régimes de vent pourrait être étudié de manière plus approfondie, les propriétés de la production régionale décrite dans cet ouvrage pourraient être corroborées et des caractéristiques pourraient être découvertes. Une telle étude pourrait aussi donner lieu à une compétition de prédiction comparable à celle effectuée pour des parcs individuels dans le cadre du projet de recherche Européen Anemos. Les résultats d'une telle compétition pourraient être analysés en profondeur afin d'augmenter notre connaissance sur le problème de la conception de modèles de prédiction de la production éolienne régionale.

Finalement, un dernier axe de recherché qui mérite d'être envisagé est la possibilité d'intégrer des mesures de variables météorologiques dans le processus de prédiction. Comme il montré dans [67], le biais des prévisions de la vitesse du vent peut être réduit en utilisant des mesures de la vitesse du vent pour filtrer les prévisions météorologiques brutes. L'étude de cette possibilité devrait être poursuivie en soi et aussi dans le cadre de la prévision de la production éolienne régionale. Comme nous l'avons montré dans cet ouvrage, une règle grossière pour la sélection des parcs de référence est de choisir les fermes qui fournissent, à priori, des informations sur la situation météorologique régnant sur l'ensemble de la région. De ce fait, il peut être intéressant d'envisager la possibilité d'intégrer des mesures de vitesse de vent prises à l'intérieur, et à l'extérieur, de la région considérée, de façon à corriger les prévisions météorologiques utilisées comme entrées du modèle de prédiction de la production. Une telle approche pourrait constituer un premier pas, statistique, vers l'amélioration de la précision des prévisions météorologiques. Un deuxième pas, possiblement plus complexe, serait le développement de modèles spécialisés capables d'exploiter de telles mesures pour améliorer directement les prévisions de production. Pour ce faire, une possibilité pourrait être de combiner des modèles physiques et statistiques de sorte qu'en obtenant une assimilation rapide de données et un temps de calcul court, des prévisions météorologiques conventionnelles puissent être améliorées, au moins pour les horizons court-terme. Une telle approche pourrait aussi, possiblement, fournir des prévisions plus précises sur les situations extrêmes, telles que les tempêtes, qui peuvent entraîner des variations massives de la puissance éolienne injectée sur les réseaux. De telles prévisions, fournies ne serait-ce que trois heures en avance de l'événement, pourraient se révéler décisives pour assurer la faisabilité économique de l'intégration à grande échelle de l'énergie éolienne dans les réseaux interconnectés.

D.3 Traduction des chapeaux de chapitre

D.3.1 Résumé Chapitre 2

Dans ce chapitre les bases de la prédiction de la production éolienne sont présentées. Une revue des modèles de prédiction existants est présentée ainsi que les critères d'évaluation des prédictions couramment utilisés. Les développements récents dans le domaine de l'estimation de l'incertitude prédiction de la production éolienne et de la valeur économique de la prédiction sont rapportés. Après ces domaines de recherche généraux, notre revue de littérature se concentre sur les modèles de prédiction de la production éolienne régionale proposés par le passé. Dans cette revue nous identifions les lacunes existantes de la prédiction de la production éolienne régionale et nous justifions les objectifs de la thèse.

D.3.2 Résumé Chapitre 3

Le but de ce chapitre est de proposer un cadre de caractérisation et d'analyse de la production éolienne régionale ainsi que de sa relation avec les variables explicatives disponibles. Ce cadre considère la production régionale comme une série temporelle simple et comme l'agrégation de plusieurs séries temporelles. De plus, le lissage statistique de la production régionale est examiné de façon détaillée. Ensuite, le lien statistique entre la production éolienne régionale et les variables issues de modèles de prévision météorologique est caractérisé en utilisant une mesure de l'information mutuelle qui existe entre les différentes variables. Dans la deuxième partie du chapitre le cadre initialement proposé est utilisé pour caractériser deux cas d'études : un comprenant 11 parcs éolien en Irlande et un comprenant la production totale de la péninsule du Jütland et de l'île de Funen, au Danemark. Les résultats de la caractérisation des deux cas sont analysés et comparés afin d'obtenir des conclusions générales sur les propriétés les plus saillantes de la production éolienne régionale.

D.3.3 Résumé Chapitre 4

Le but de ce chapitre est d'examiner les modèles de prévision de la production éolienne régionale et de fournir aux prévisionnistes des guides pour la mise en oeuvre de tels modèles. Dans ce but, le problème de la prédiction de la production éolienne régionale est abordé comme un problème d'apprentissage statistique. De ce point de vue, les modèles de prédiction existants sont étudiés et trois approches de combinaison de modèles de base (ou sous-modèles) sont identifiées. Nous concentrons notre étude ensuite sur la caractérisation de la performance obtenue avec différents sous-modèles dans les trois approches de combinaison. Deux modèles de base sont utilisés dans cette caractérisation : un de réseau de neurones floues (F-NN) de l'état de l'art et un nouveau modèle proposé ici, baptisé Regressive Power Curve Model (RPC). Les performances de ce dernier sont tout d'abord validées en les comparant à celles de plusieurs autres modèles sur trois jeux de données de l'état de l'art, jeux de données dont les complexités des terrains sont différentes. Les deux sous-modèles sont ensuite utilisés pour étudier les performances relatives des trois approches de combinaison et pour examiner l'influence des sous-modèles sur les performances obtenues. Cette étude est menée sur les deux cas d'études décrits dans le chapitre précédent.

D.3.4 Résumé Chapitre 5

Ce chapitre examine l'impact de la sélection de variables explicatives sur la précision de prédiction. Dans la première partie, nous proposons une méthodologie pour l'étude de l'influence de différents sous-ensembles de variables explicatives sur la précision des prédictions de la production éolienne régionale. Cette méthodologie est ensuite appliquée aux séries temporelles de mesures de production et aux prévisions de vitesses de vent issues de modèles météorologiques. A partir des résultats de cette étude, des conclusions générales sont obtenues et le besoin d'effectuer une sélection de variables explicatives dans le cadre de la prédiction de la production éolienne régionale est clairement identifié. Cette étude mène à la deuxième partie du chapitre qui examine l'application de méthodes sélection de variables explicatives au problème de la prédiction de la production éolienne régionale. Cette partie débute par une revue des méthodologies de sélection de variables proposées dans la littérature. A partir de cette revue trois méthodes de sélection sont proposées, chacune se basant sur une méthodologie différente. La première méthode est une méthode de l'état de l'art qui nous sert de référence. Les deux autres méthodes : une méthode filtre basée sur l'agrégation de variables et une méthode wrapper basée sur la stabilité de transition entre des sous-ensembles de variables, sont des méthodes originales. Ces dernières exploitent certaines caractéristiques du problème de la prédiction de la production éolienne régionale. Les trois méthodes sont comparées dans les cas d'études Irlandais et Danois. A partir de cette comparaison, les performances de méthodes sont analysées et des conclusions sur leur utilité sont obtenues.

Bibliography

[1] International Panel on Climate Change, *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, J. Houghton, Y. Ding, D. Griggs, M. Noguer, P. van der Linden, X. Dai, K. Maskell, and C. Johnson, Eds. Cambridge, United Kingdom and New York, NY, USA,: Cambridge University Press, 2001. [Online]. Available: http://www.grida.no/climate/ipcc_tar/wg1/index.htm

[2] International Energy Agency, *World Energy Outlook 2004*. IEA PUBLICATIONS, 2004. [Online]. Available: http://www.iea.org/textbase/nppdf/free/2004/weo2004.pdf

[3] ——, "The global renewable energy policies and measures database," Website, June 2006. [Online]. Available: http://renewables.iea.org

[4] "Directive 2001/77/ec of the european parliament and of the council of 27 september 2001 on the promotion of electricity produced from renewable energy sources in the internal electricity market," *Official Journal of the European Commission*, vol. L 283, pp. 33–40, 2001. [Online]. Available: http://europa.eu.int/eur-lex/pri/en/oj/dat/2001/l_283/l_28320011027en00330040.pdf

[5] L. Bird, M. Bolinger, T. Gagliano, R. Wiser, M. Brown, and B. Parsons, "Policies and market factors driving wind power development in the united states," *Energy Policy*, vol. 33, no. 11, pp. 1397–1407, July 2005. [Online]. Available: http://www.sciencedirect.com/science/article/B6V2W-4BRSFC9-1/2/d7c7da67036a7bc64f146248116d34f4

[6] Ministry of Non-Conventional Energy Sources Government of India, *Annual Report 2005-2006*, 2006. [Online]. Available: http://mnes.nic.in/frame.htm?majorprog.htm

[7] Global Wind Energy Council, "Global wind 2006 report," Global Wind Energy Council, Brussels, Belgium, Tech. Rep., 2006. [Online]. Available: http://www.gwec.net/fileadmin/documents/Publications/gwec-2006_final_01.pdf

[8] M. Hoogwijk, B. de Vries, and W. Turkenburg, "Assessment of the global and regional geographical, technical and economic potential of onshore wind energy," *Energy Economics*, vol. 26, no. 5, pp. 889–919, Sept. 2004. [Online]. Available: http://www.sciencedirect.com/science/article/B6V7G-4CX6XY1-1/2/4233dcd7ca19dd51b566a0438aabc1d4

[9] Global Wind Energy Council, “Global wind 2005 report,” Global Wind Energy Council, Brussels, Belgium, Tech. Rep., 2005. [Online]. Available: http://www.gwec.net/fileadmin/documents/Publications/Global_WindPower_05_Report.pdf

[10] P. A. Østergaard, “Ancillary services and the integration of substantial quantities of wind power,” *Applied Energy*, vol. 83, no. 5, pp. 451–463, May 2006. [Online]. Available: http://www.sciencedirect.com/science/article/B6V1T-4GKWJ77-2/2/e30278d384303f8de323e06d2fe2027b

[11] D. Anderson and M. Leach, “Harvesting and redistributing renewable energy: on the role of gas and electricity grids to overcome intermittency through the generation and storage of hydrogen,” *Energy Policy*, vol. 32, no. 14, pp. 1603–1614, Sept. 2004. [Online]. Available: http://www.sciencedirect.com/science/article/B6V2W-49CN3VC-1/2/9eb357d90ddba2736e32bd4f664b864d

[12] F. Graves, T. Jenkin, and D. Murphy, “Opportunities for electricity storage in deregulating markets,” *The Electricity Journal*, vol. 12, no. 8, pp. 46–56, Oct. 1999. [Online]. Available: http://www.sciencedirect.com/science/article/B6VSS-3XV2K6H-6/2/9600585ac5037a6bf7d974010576e1dd

[13] P. Denholm, “Improving the technical, environmental and social performance of wind energy systems using biomass-based energy storage,” *Renewable Energy*, vol. 31, no. 9, pp. 1355–1370, July 2006. [Online]. Available: http://www.sciencedirect.com/science/article/B6V4S-4GY870M-1/2/c32ddbf68374596fa6817f4986c7056d

[14] E. Castronuovo and J. Lopes, “On the optimization of the daily operation of a wind-hydro power plant,” *Power Systems, IEEE Transactions on*, vol. 19, no. 3, pp. 1599–1606, 2004.

[15] R. Castro and L. Ferreira, “A comparison between chronological and probabilistic methods to estimate wind power capacity credit,” *Power Systems, IEEE Transactions on*, vol. 16, no. 4, pp. 904–909, 2001.

[16] M. Grubb, “Value of variable sources on power systems,” *Generation, Transmission and Distribution [see also IEE Proceedings-Generation, Transmission and Distribution], IEE Proceedings C*, vol. 138, no. 2, pp. 149–165, 1991.

[17] P. Eriksen, T. Ackermann, H. Abildgaard, P. Smith, W. Winter, and J. Rodriguez Garcia, “System operation with high wind penetration,” *Power and Energy Magazine, IEEE*, vol. 3, no. 6, pp. 65–74, 2005.

[18] E. DeMeo, W. Grant, M. Milligan, and M. Schuerger, “Wind plant integration [wind power plants],” *Power and Energy Magazine, IEEE*, vol. 3, no. 6, pp. 38–46, 2005.

[19] E. Fagan, S. Grimes, J. McArdle, P. Smith, and M. Stronge, “Grid code provisions for wind generators in ireland,” in *Power Engineering Society General Meeting, 2005. IEEE*, 2005, pp. 1241–1247 Vol. 2.

[20] RTE, “Bilan prévisionel de l’équilibre offre demande d’électricité en France,” Gestionnaire du Réseau de Transport d’Electricité, Tech. Rep., 2005.

[21] S.-V. Thor and P. Weis-Taylor, “Long-term research and development needs for wind energy for the time frame 2000-2020,” *Wind Energy*, vol. 5, no. 1, pp. 73–75, 2002. [Online]. Available: http://dx.doi.org/10.1002/we.65

[22] R. Kimura, “Numerical weather prediction,” *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 90, no. 12-15, pp. 1403–1414, Dec. 2002. [Online]. Available: http://www.sciencedirect.com/science/article/B6V3M-476JXKB-6/2/a1fa97cbfb73070a1494354df93328fa

[23] B. F. Hobbs, M. H. Rothkopf, R. P. O'Neil, and H.-p. Chao, Eds., *The Next Generation of Electric Power Unit Commitment Models*, ser. International Series in Operations Research & Management Science. Dordrecht: Kluwer Academic Publishers, 2001.

[24] U. Focken, M. Lange, K. Mönnich, H. G. Waldl, H.-P.andBeyer, and A. Luig, "Short-term prediction of the aggregated power output of wind farms - a statistical analysis of the reduction of the prediction error by spatial smoothing effects," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 90, no. 3, pp. 231–246, Mar. 2002. [Online]. Available: http://www.sciencedirect.com/science/article/B6V3M-44Y0X3R-1/2/e22ae2cbfc04e5d029065a3ad14b20fe

[25] G. Kariniotakis, G. S. Stavrakakis, and E. Nogaret, "Wind power forecasting using advanced neural networks models," *IEEE Transactions on Energy Conversion*, vol. 11, no. 4, pp. 762–767, Dec. 1996.

[26] H. Madsen, P. Pinson, G. Kariniotakis, H. A. Nielsen, and T. S. Nielsen, "Standardizing the performance evaluation of shortterm wind power prediction models," *Wind Engineering*, vol. 29, no. 6, pp. 475–485, Dec. 2005.

[27] G. Kariniotakis, P. Pinson, N. Siebert, G. Giebel, and R. Barthelmie, "The state of the art in short-term prediction of wind power - from an offshore perspective," in *Proceeding of the 2004 SeaTechWeek*, Brest, France, 20-21 October 2004. [Online]. Available: http://anemos.cma.fr/modules.php?name=News&file=article&sid=17

[28] G. Giebel, G. Kariniotakis, and R. Brownsword, "The state-of-the-art in short-term prediction of wind power - from a danish perspective," in *Proceedings of the 4th International Workshop on Large-Scale Integration of Wind Power and Transmission Networks for Offshore Wind Farms*, Billund, Denmark, 21-23 October 2003.

[29] S. Heier, *Grid integration of wind energy conversion systems*. Chichester ; New York: Wiley, 1998.

[30] G. E. P. Box and G. M. Jenkins, *Time Series Analysis forecasting and control*, revised ed., ser. Holden-Day series in Time Series Analysis and Digital Processing, R. Enders, Ed. Oakland, California, USA: Holden-Day, 1976.

[31] M. Milligan, M. N. Schwartz, and Y. Wan, "Statistical wind power forecasting for u.s. wind farms," National Renewable Energy Laboratory, 1617 Cole Boulevard, Golden, Colorado, Research Report NREL/CP-500-35087, Nov. 2003.

[32] G. Kariniotakis, I. Martí, D. Casas, P. Pinson, T. S. Nielsen, H. Madsen, J. Giebel, G. andUsaola, I. Sanchez, A. M. Palomares, R. Brownsword, J. Tambke, U. Focken, P. Lange, M. andLouka, G. Kallos, C. Lac, G. Sideratos, and G. Descombes, "What performance can be expected by short-term wind power prediction models depending on site characteristics?" in *Proceedings of the 2004 European Wind Energy Association Conference EWEC'04*, London, UK., Nov. 2004. [Online]. Available: http://www.2004ewec.info/files/23_1400_georgekariniotakis_01.pdf

[33] C. Chatfield, *Time-series forecasting*. Boca Raton: Chapman & Hall/CRC, 2000.

[34] H. Madsen, G. Kariniotakis, H. A. Nielsen, T. S. Nielsen, and P. Pinson, "A protocol for standardising the performance evaluation of short-term wind power prediction models," in *Proceedings of the Global WindPower 2004 Conference*, Chicago, Illinois, Mar. 2004.

[35] S. Lang, J. Möhrlen, J. Jørgensen, B. O. Gallachóir, and E. McKeogh, "Application of a multi-scheme ensemble prediction system for wind power forecasting in ireland and comparison with validation results from denmark and germany," in *Proceedings of the 2006 European Wind Energy Conference, EWEC'06*, Athens, Greece, 2006. [Online]. Available: http://www.ewec2006proceedings.info/

[36] P. Pinson, "Estimation of the uncertainty in wind power forecasting," Ph.D. dissertation, Ecole des Mines de Paris, Center for Energy and Processes, Rue Claude Daunesse, Sophia Antipolis, 2006.

[37] T. S. Nielsen, A. Joensen, H. Madsen, L. Landberg, and G. Giebel, "A new reference for wind power forecasting," *Wind Energy*, vol. 1, no. 1, pp. 29–34, 1998. [Online]. Available: http://dx.doi.org/10.1002/(SICI)1099-1824(199809)1:1<29::AID-WE10>3.0.CO;2-B

[38] P. Pinson, N. Siebert, and G. Kariniotakis, "Forecasting of regional wind generation by a dynamic fuzzy-neural networks based upscaling approach," in *Proceedings of the 2003 European Wind Energy Association Conference EWEC'03*, Madrid, Spain, 16-19 July 2003.

[39] E. Nogaret, G. Stavrakakis, J. Bonin, G. Kariniotakis, B. Papadias, G. Contaxis, M. Papadopoulos, N. Hatziargyriou, S. Papathanassiou, J. Garopoulos, E. Karagounis, J. Halliday, G. Dutton, J. Pedas-Lopes, A. Androutsos, and P. Pligoropoulos, "Development and implementation of an advanced control system for medium size wind-diesel systems." in *Proceedings of the 1994 European Wind Energy Association Conference EWEC'94*, Thesaloniki, Greece, Oct. 1994, pp. 599–604.

[40] S. Makridakis, S. C. Wheelwright, and V. E. McGee, *Forecasting : methods and applications*, 2nd ed. John Wiley & Sons, 1983.

[41] H. Fukuda, S. Tamaki, M. Nakamura, H. Nagai, F. Shijo, S. Asato, and K. Onaga, "The development of a wind velocity prediction method based on a data mining type auto-regressive model," in *Proceedings of the 2001 European Wind Energy Association Conference EWECŠ01*, Copenhagen, Denmark, July 2001, pp. 741–744.

[42] M. J. Durán, D. Cros, and J. Riquelme, "Short-term wind power prediction based on ar models," in *Proceedings of the 9th Spanish-Portuguese Congress on Electrical Engineering*, 2005. [Online]. Available: http://www.aedie.org/9CHLIE-paper-send/374-DURAN.pdf

[43] A. Dutton, G. Kariniotakis, J. Halliday, and E. Nogaret, "Load and wind power forecasting methods for the optimal management of isolated power systems with high wind penetration," *Wind Engineering*, vol. 23, no. 2, pp. 69–87, 1999.

[44] M. C. Alexiadis, P. S. Dokopoulos, and H. S. Sahsamanoglou, "Wind speed and power forecasting based on spatial correlation models," *IEEE Transactions on Energy Conversion*, vol. 14, no. 3, pp. 836–842, Sept. 1999.

[45] S. Li, D. C. Wunsch, E. A. OŠHair, and M. G. Giesselmann, "Using neural networks to estimate wind turbine power generation," *IEEE Transactions on Energy Conversion*, vol. 16, no. 3, pp. 276–282, Sept. 2001.

[46] G. Kariniotakis, "Contribution au développement d'un système de contrôle avancé pour les systèmes éolien-diesel autonomes," Ph.D. dissertation, Ecole des Mines de Paris, Center for Energy Studies, Rue Claude Daunesse, Sophia Antipolis, Dec. 1996.

[47] A. Sfetsos, "Time series forecasting of wind speed and solar radiation for renewable energy sources," Ph.D. dissertation, Imperial College, London, 1999.

[48] ——, "A comparison of various forecasting techniques applied to mean hourly wind speed time series," *Renewable Energy*, vol. 21, no. 1, pp. 23–35, Sept. 2000.

[49] S. Enomoto, N. Inomata, T. Yamada, H. Chiba, R. Tanikawa, T. Oota, and H. Fukuda, "Prediction of power output from wind farm using local meteorological analysis," in *Proceedings of the 2001 European Wind Energy Conference, EWEC'01*, Copenhagen, Denmark, July 2001, pp. 822–825.

[50] M. Magnusson and L. Wern, "Wind energy predictions using cfd and hirlam forecasts," in *Proceedings of the 2001 European Wind Energy Conference, EWEC'01*, Copenhagen, Denmark, July 2001, pp. 861–863.

[51] J.-M. Noël and R. Chevallaz-Perrier, "Wind resource assessment with ariawind in a complex site," in *Proceedings of the 2001 European Wind Energy Conference, EWEC'01*, Copenhagen, Denmark, July 2001, pp. 872–874.

[52] I. Troen and E. L. Petersen, *European Wind Atlas.* Roskilde, Denmark: Risø National Laboratory, 1989.

[53] L. Landberg, "Short-term prediction of local wind conditions," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 89, no. 3-4, pp. 235–245, Mar. 2001. [Online]. Available: http://www.sciencedirect.com/science/article/B6V3M-42DP3K0-3/2/1d5079d0f591faa0b5fe22392089d20b

[54] R. Watson, L. Landberg, R. Costello, D. McCoy, and P. OŠDonnell, "Evaluation of the prediktor wind power forecasting system in ireland," in *Proceedings of the 2001 European Wind Energy Conference, EWEC'01*, Copenhagen, Denmark, July 2001, pp. 861–863.

[55] G. Giebel, L. Landberg, T. S. Nielsen, and H. Madsen, "The zephyr-project : The next generation prediction system," in *Proceedings of the 2001 European Wind Energy Conference, EWEC'01*, Copenhagen, Denmark, July 2001, pp. 777–780.

[56] H. G. Beyer, D. Heinemann, H. Mellinghoff, K. Mönnich, and H.-P. Waldl, "Forecast of regional wind power output of wind turbines," in *Proceedings of the 1999 European Wind Energy Conference, EWEC'99*, Nice, France, Mar. 1999, pp. 1070–1073.

[57] U. Focken, M. Lange, and H.-P. Waldl, "Previento: a wind power prediction system with an innovative upscaling algorithm," in *Proceedings of the 2001 European Wind Energy Association Conference EWEC'01*, Copenhagen, Denmark, 2-6 July 2001, pp. 826–829.

[58] I. Martí, D. Cabezón, J. Villanueva, M. J. Sanisidro, Y. Loureiro, E. Cantero, and J. Sanz, "Localpred and Regiopred. Advanced tools for wind energy prediction in complex terrain," in *Proceedings of the European Wind Energy Conference EWEC'03*, Madrid, Spain, 16-19 June 2003.

[59] B. Bailey, M. C. Brower, and J. Zack, "Short-term wind forecasting Ű development and application of a mesoscale model," in *Proceedings of the 1999 European Wind Energy Conference, EWEC'99*, Nice, France, Mar. 1999, pp. 1062–1065.

[60] T. S. Nielsen, H. Madsen, and J. Tøfting, "Experiences with statistical methods for wind power prediction," in *Proceedings of the 1999 European Wind Energy Conference, EWEC'99*, Nice, France, Mar. 1999, pp. 1066–1069.

[61] T. S. Nielsen and H. Madsen, "Statistical methods for predicting wind power," in *Proceedings of the 1997 European Wind Energy Conference, EWEC'97*, Dublin, Ireland, Oct. 1997, pp. 755–758.

[62] L.-X. Wang, *Adaptive fuzzy systems and control : design and stability analysis.* Englewood Cliffs, N.J.: PTR Prentice Hall, 1994, li-Xin Wang.; Includes bibliographical references (p. 221-227) and index.

[63] G. Kariniotakis, E. Nogaret, A. Dutton, J. Halliday, and A. Androutsos, "Evaluation of advanced wind power and load forecasting methods for the optimal management of isolated power systems," in *Proceedings of the 1999 European Wind Energy Conference, EWEC'99*, Nice, France, Mar. 1999, pp. 1082–1085.

[64] G. Kariniotakis and D. Mayer, "An advanced on-line wind resource prediction system for the optimal management of wind parks," in *3rd MedPower Conference*, Athens, Greece, Nov. 2002.

[65] I. Sanchez and J. Usaola, “Sipreólico - a wind power prediction system based on flexible combination of dynamic models. application to the spanish power system,” in *Proceedings of the 2002 World Wind Energy Conference*, Berlin, Germany, June 2002.

[66] J. Parkes and A. Tindal, “Forecasting short term wind farm production in complex terrain,” in *Proceedings of the 2004 European Wind Energy Conference, EWEC’04*, London, U.K., Nov. 2004. [Online]. Available: http://www.2004ewec.info/files/23_1400_jeremyparkes_01.pdf

[67] P. Louka, G. Galanis, N. Siebert, G. Kariniotakis, P. Katsafados, G. Kallos, and I. Pytharoulis, “Improvements in wind speed forecasts for wind power prediction purposes using kalman filtering,” in *CD-Rom Proceedings of the CMMSE 2005*, Alicante, Spain, June 2005.

[68] M. Lange, “On the uncertainty of wind power predictions. analysis of the forecast accuracy and statistical distribution of errors,” *Journal of Solar Energy Engineering- Transactioncs of the ASME*, vol. 127, no. 2, pp. 177–184, May 2005.

[69] A. Luig, S. Bofinger, and H. Beyer, “Analysis of confidence intervals for the prediction of regional wind power output,” in *Proceedings of the 2001 European Wind Energy Conference, EWEC’01*, Copenhagen, Denmark, June 2001, pp. 725–728.

[70] H. A. Nielsen, T. S. Nielsen, and H. Madsen, “Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts,” in *Proceedings of the 2004 European Wind Energy Conference, EWEC’04, Scientific Track*, London, United Kingdom„ Nov. 2004, pp. 34Ű–38.

[71] J. B. Bremnes, “Probabilistic wind power forecasts using local quantile regression,” *Wind Energy*, vol. 7, no. 1, pp. 47–54, 2004. [Online]. Available: http://dx.doi.org/10.1002/we.107

[72] P. Pinson and G. Kariniotakis, “On-line assessment of prediction risk for wind power production forecasts,” in *Proceedings of the 2003 European Wind Energy Conference, EWEC’03*, Madrid, Spain, June 2003.

[73] G. Giebel, J. Badger, L. Landberg, H. A. Nielsen, T. S. Nielsen, H. Madsen, H. Sattler, K.andFeddersen, H. Vedel, J. Tøfting, L. Kruse, and L. Voulund, “Wind power prediction using ensembles,” Risø National Laboratory, Roskilde, Denmark, Research Report Risø-R-1527(EN), Sept. 2005. [Online]. Available: http://www.risoe.dk/rispubl/VEA/veapdf/ris-r-1527.pdf

[74] H. A. Nielsen, H. Nielsen, T. S.and Madsen, and K. Sattler, “Wind power ensemble forecasting,” in *CD-Proceedings of the 2004 Global Windpower Conference*, Chicago, Illinois (USA), Mar. 2004.

[75] S. Haanpää, S. Lehtonen, L. Peltonen, and E. Talockaite, *Impacts of winter storm Gudrun of 7th Ű 9th January 2005 and measures taken in Baltic Sea Region.* Astra Project, 2006, report. [Online]. Available: http://www.astra-project.org/cms/sites/download/ASTRA_WSS_report_final.pdf

[76] M. R. Milligan, A. H. Miller, and F. Chapman, “Estimating the economic value of wind forecasting to utilities,” National Renewable Energy Laboratory, Golden, Colorado (U.S.), Research Report NREL/TP-441-7803, May 1995. [Online]. Available: http://www.nrel.gov/docs/legosti/old/7803.pdf

[77] B. Sørensen and P. Meibom, “Can wind power be sold in a deregulated electricity market?” in *Proceedings of the 1999 European Wind Energy Conference, EWEC’99*, Nice, France, Mar. 1999, pp. 375–378.

[78] A. Fabbri, T. Gomez San Roman, J. Rivier Abbad, and V. Mendez Quezada, “Assessment of the cost associated with wind generation prediction errors in a liberalized electricity market,” *Power Systems, IEEE Transactions on*, vol. 20, no. 3, pp. 1440–1446, 2005.

[79] S. Bofinger, A. Luig, and H. G. Beyer, “Qualification of wind power forecasts,” in *CD-Proceedings of the Global Wind Power Conference*, Paris, France, Apr. 2002.

[80] P. Pinson, C. Chevallier, and G. Kariniotakis, “Optimizing benefits from wind power participation in electricity markets using advanced tools for wind power forecasting and uncertainty assessment,” in *Proceedings of the 2004 European Wind Energy Conference, EWEC’04*, London, U.K., Nov. 2004. [Online]. Available: http://www.2004ewec.info/files/23_1400_pierrepinson_01.pdf

[81] P. Pinson, C. Chevallier, and N. Kariniotakis, George, “Trading wind generation with short-term probabilistic forecasts of wind power,” *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 1148–1156, Aug. 2007.

[82] P. Pinson and G. Kariniotakis, “On-line adaptation of confidence intervals based on weather stability for wind power forecasting,” in *CD-Proceeding of the 2004 Global Wind Windpower Conference*, Chicago, Illinois (U.S.), Mar. 2004.

[83] U. Focken, M. Lange, and H.-P. Waldl, “Reduction of wind power prediction error by spatial smoothing effects,” in *Proceedings of the 2001 European Wind Energy Association Conference EWEC’01*, Copenhagen, Denmark, 2-6 July 2001, pp. 822–825.

[84] T. S. Nielsen, H. Madsen, H. A. Nielsen, L. Landberg, and G. Giebel, “Prediction of regional wind power,” in *Proceeding of the 2002 Global Windpower Conference*, Paris, France, April 2002.

[85] B. Ernst, K. Rohrig, P. Schorn, and H. Regber, “Managing 3000 mw wind power in a transmission system operation centres,” in *Proceedings of the 2001 European Wind Energy Association Conference EWEC’01*, Copenhagen, Denmark, 2-6 July 2001, pp. 890–893.

[86] K. Rohrig, B. Ernst, M. Hoppe-Kilpper, and F. Schlögl, “Online-monitoring and prediction of wind power in german transmission system operation centres,” in *Proceeding WWEC 2003 - World Wind Energy Conference*, Cape Town, South Africa, November 2003. [Online]. Available: http://www.iset.uni-kassel.de/abt/FB-I/publication/03-11-01_WWEC2003.pdf

[87] S. Lang, J. Möhrlen, J. Jørgensen, B. O. Gallachóir, and E. McKeogh, “Aggregate forecasting of wind generation on the irish grid using a multi-scheme ensemble prediction system,” in *Proceedings, Renewable Energy in Maritime Island Climates, 2nd Conference*, Dublin, Ireland, April 2006.

[88] M. Lange and U. Focken, “State-of-the-art in wind power prediction in germany and international developments,” in *Proceedings of the Second Workshop of International Feed-In Cooperation*, Berlin, Dec. 2005. [Online]. Available: http://www.energymeteo.de/de/media/fic_eeg_article.pdf

[89] A. Yamaguchi, T. Ishihara, K. Sakai, T. Ogawa, and Y. Fujino, “A physical-statistical approach for the regional power forecasting,” in *Proceedings of the 2007 European Wind Energy Conference, EWEC’07*, Milano, Italy, May 2007.

[90] T. Ishihara, A. Yamaguchi, T. Ogawa, K. Sakai, and Y. Fujino, “An upscaling approach for the regional wind power forecasting,” in *Proceedings of the 2007 European Wind Energy Conference, EWEC’07*, Milano, Italy, May 2007.

[91] T. S. Nielsen, “Online prediction and control in nonlinear stochastic systems,” Ph.D. dissertation, Department of Mathematical Modelling, Technical University of Denmark, 2002.

[92] T. S. Nielsen and H. A. Nielsen, “Using meteorological forecasts in on-line predictions of wind power. chapter 5: Estimation methods,” 1999, iSBN: 87-90707-18-4. [Online]. Available: http://www2.imm.dtu.dk/pubdb/p.php?1706

[93] L. v. Bremen, N. Saleck, and D. Heinemann, "Enhanced regional forecasting considering single wind farm distribution for upscaling," *Journal of Physics: Conference Series*, vol. 75, Aug. 2007. [Online]. Available: http://stacks.iop.org/1742-6596/75/012040

[94] G. Giebel, "On the benefits of distributed generation of wind energy in europe," Ph.D. dissertation, Carl von Ossietzky Universität, Oldenburg, Germany, 2000.

[95] "London array project website," Aug. 2006. [Online]. Available: http://www.londonarray.com

[96] A. Sfetsos and C. Siriopoulos, "Time series forecasting of averaged data with efficient use of information," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 35, no. 5, pp. 738–745, Sept. 2005.

[97] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed., ser. Electrical & Electronic Engineering Series. New York, NY, USA: McGraw-Hill, 1991.

[98] L. Balea, N. Siebert, G. Kariniotakis, and E. Peirano, "Quantification of capacity credit and reserve requirements from the large-scale integration of wind energy in the french power system," in *CD-Proceedings of the 2004 Global Windpower Conference*, Chicago, Illinois (U.S.), Mar. 2004.

[99] R. M. Gray, *Entropy and Information Theory*. New York, NY, USA: Springer-Verlag, 1990.

[100] D. W. Scott, *Multivariate Density Estimation, Theory, Practice, and Visialization*, ser. Probability and Mathematical Statistics. New York: Wiley-Interscience, 1992, iSBN: 0-471-54770-0.

[101] B. Bonnlander, "Nonparametric selection of input variables for connectionist learning," Ph.D. dissertation, University of Colorado Department of Computer Science, 1996.

[102] European Wind Energy Association, "European capacity map 2006," 2006. [Online]. Available: http://www.ewea.org/fileadmin/ewea_documents/documents/publications/statistics/070129_Wind_map_2006.pdf

[103] V. Miranda, C. Cerqueira, and C. Monteiro, "Training a fis with epso under an entropy criterion for wind power prediction," in *On-line proceedings of the 9th International Conference on Probabilistic Methods Applied to Power Systems*, Stockholm, Sweden, June 2006. [Online]. Available: http://www.congrex.se/pmaps/pdf/1.3/1.3_miranda_v.pdf

[104] H. Holttinen, P. Saarikivi, S. Repo, J. Ikäheimo, and G. Koreneff, "Prediction error and balancing costs for wind power production in finland," in *Proceedings of the Sixth International Workshop on Large-Scale Integration of Wind Power and Transmission Networks for Offshore Wind Famrs*, R. L. Hendriks, B. C. Ummels, and T. Ackermann, Eds., Delft, The Netherlands, Oct. 2006, pp. 113–120.

[105] A. Woyte, G. Palmers, and L. Dewilde, "Balancing costs for large wind farms: The influence of short-term forecasting and the regulatory framework," in *Proceedings of the Sixth International Workshop on Large-Scale Integration of Wind Power and Transmission Networks for Offshore Wind Famrs*, R. L. Hendriks, B. C. Ummels, and T. Ackermann, Eds., Delft, The Netherlands, Oct. 2006, pp. 121–126.

[106] V. N. Vapnik, *The nature of statistical learning theory*, 2nd ed., ser. Statistics for Engineering and Information Science. New York, NY, USA: Springer-Verlag New York, Inc., 1999.

[107] F. Rosenblatt, *Principles of neurodynamics: Perceptron and theory of brain mechanisms*. Washington D.C: Spartan Books, 1962.

[108] V. N. Vapnik, *Estimation of Dependencies Based on Empirical Data*. New York: Springer, 1982.

[109] V. N. Vapnik and A. J. Chervonenkis, "The necessary and sufficient conditions for consistency of the method of empirical risk minimization," *Pattern Recognition and Image Analysis*, vol. 1, no. 3, pp. 284–305, 1991.

[110] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *COLT '92: Proceedings of the fifth annual workshop on Computational learning theory*. New York, NY, USA: ACM Press, 1992, pp. 144–152.

[111] Y. LeCun, "Une procédure d'apprentissage pour réseau a seuil asymmetrique (a learning scheme for asymmetric threshold networks)," in *Proceedings of Cognitiva 85*, Paris, France, 1985, pp. 599–604. [Online]. Available: http://yann.lecun.com/exdb/publis/index.html

[112] D. B. Parker, "Learning-logic," MIT Center for Computational Research in Economics and Management Science, Cambridge, MA, TR-47, 1985.

[113] L. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, June 1965. [Online]. Available: http://www.sciencedirect.com/science/article/B7MFM-4DX43MN-W3/2/f244f7a33f31015e819042700cd83047

[114] V. Kecman and B.-M. Pfeiffer, "Exploiting the structural equivalence of learning fuzzy systems and radial basis function neural networks," in *Proceedings of the Second European Congress on Intelligent Techniques and Soft Computing EUFIT-94, Aachen, Gemany*, vol. 1, 1994, pp. 58–66.

[115] S. Makridakis and M. Hibon, "The m3-competition: results, conclusions and implications," *International Journal of Forecasting*, vol. 16, no. 4, pp. 451–476, 2000. [Online]. Available: http://www.sciencedirect.com/science/article/B6V92-41J6944-3/2/77108ea44b47436e4527d06abc3a8ac5

[116] A. J. Koning, P. H. Franses, M. Hibon, and H. Stekler, "The m3 competition: Statistical tests of the results," *International Journal of Forecasting*, vol. 21, no. 3, pp. 397–409, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/B6V92-4F011CS-1/2/53b7026849c6c1659a592a5451b122eb

[117] R. Wang, V. Storey, and C. Firth, "A framework for analysis of data quality research," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 7, no. 4, pp. 623–640, 1995.

[118] Y. Wand and R. Y. Wang, "Anchoring data quality dimensions in ontological foundations," *Commun. ACM*, vol. 39, no. 11, pp. 86–95, 1996.

[119] R. E. Bellman, *Adaptive Control Processes: A Guided Tour*. Princeton, NJ: Princeton Universsity Press, 1961.

[120] P. L. Meyer, *Probabilidade: Aplicações à Estatística*, 2nd ed. Rio de Janeiro, Brasil: Livros Técnicos e Cientìficos Editora, 1983.

[121] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, no. 1, pp. 81–106, Mar. 1986. [Online]. Available: http://dx.doi.org/10.1007/BF00116251

[122] P. Langley, W. Iba, and K. Thompson, "An analysis of bayesian classifiers," in *Proceedings of the 10th National Conference on Artificial Intelligence*, W. R. Swartout, Ed. San Jose, CA: The AAAI Press / The MIT Press, July 12-16 1992, pp. 223–228.

[123] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Machine Learning*, vol. 29, no. 2, pp. 131–163, Nov. 1997. [Online]. Available: http://dx.doi.org/10.1023/A:1007465528199

[124] T. S. Nielsen, H. Madsen, H. A. Nielsen, P. Pinson, G. Kariniotakis, N. Siebert, I. Martí, M. Lange, U. Focken, L. von Bremen, P. Louka, G. Kallos, and G. Galanis, "Short-term wind power forecasting using advanced statistical methods," in *Proceedings of the 2006 European Wind Energy Conference, EWEC'06*, Athens, Greece, 2006. [Online]. Available: http://www.ewec2006proceedings.info/index.php

[125] I. Sánchez, "Short-term prediction of wind energy production," *International Journal of Forecasting*, vol. 22, no. 1, pp. 43–56, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/B6V92-4HDP6VM-1/2/0f5ac6ae7950992b63e41d2dfaa2d49a

[126] C. Peck, A. Dhawan, and C. Meyer, "Genetic algorithm based input selection for a neural network function approximator with applications to ssme health monitoring," in *IEEE International Conference on Neural Networks, 1993.*, vol. 2, San Francisco, CA, 1993, pp. 1115–1122.

[127] T. Chow and D. Huang, "Estimating optimal feature subsets using efficient estimation of high-dimensional mutual information," *Neural Networks, IEEE Transactions on*, vol. 16, no. 1, pp. 213–224, 2005.

[128] D. Paravan, T. Stokelj, and R. Golob, "Selecting input variables for hpp reservoir water inflow forecasting using mutual information," in *Proceedings of the 2001 IEEE Porto Power Tech Conference*, vol. 2, Porto, Portugal, Sept. 2001.

[129] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, Mar. 2003.

[130] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, no. 1-2, pp. 273–324, Dec. 1997.

[131] A. Blum and R. L. Rivest, "Training a 3-node neural network is NP-complete," *Neural Networks*, vol. 5, no. 1, pp. 117–127, 1992.

[132] D. M. Chickering, D. Geiger, and D. Heckerman, "Learning Bayesian Networks is NP-Hard," Microsoft Research, Advanced Technology Division, Microsoft Corporation, One Microsoft Way, Redmond, WA 98052, Tech. Rep. MSR-TR-94-17, Nov. 1994.

[133] A. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artificial Intelligence*, vol. 97, no. 1-2, pp. 245–271, Dec. 1997.

[134] M. Hall, "Correlation-based feature selection for machine learning," Ph.D. dissertation, Waikato University, Department of Computer Science., Hamilton, NZ, Apr. 1999. [Online]. Available: citeseer.ist.psu.edu/hall99correlationbased.html

[135] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *Neural Networks, IEEE Transactions on*, vol. 5, no. 4, pp. 537–550, 1994.

[136] D. A. Bell and H. Wang, "A formalism for relevance and its application in feature subset selection," *Machine Learning*, vol. 41, no. 2, pp. 175–195, Nov. 2000. [Online]. Available: http://dx.doi.org/10.1023/A:1007612503587

[137] K. Li, "A new input selection method for neural modeling of nonlinear complex systems," in *Intelligent Control and Automation, 2004. WCICA 2004. Fifth World Congress on*, vol. 3, 2004, pp. 2018–2021 Vol.3.

[138] P. Mitra, C. Murthy, and S. Pal, "Unsupervised feature selection using feature similarity," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 3, pp. 301–312, 2002.

[139] R. Šindelář and R. Babuška, "Input selection for nonlinear regression models," *IEEE Transactions on Fuzzy Systems*, vol. 12, no. 5, pp. 688–696, Oct. 2004.

[140] B. V. Bonnlander and A. S. Weigend, "Selecting input variables using mutual information and nonparametric density estimation," in *Proc. of the 1994 Int. Symp. on Artificial Neural Networks (ISANN"94)*, Tainan, Taiwan, 1994, pp. 42–50. [Online]. Available: citeseer.ist.psu.edu/bonnlander96selecting.html

[141] K. Torkkola, "Feature extraction by non-parametric mutual information," *Journal of Machine Learning Research*, vol. 3, pp. 1415–1438, Mar. 2003.

[142] A. Singhal and D. Seborg, "Clustering of multivariate time-series data," in *Proceedings of the 2002 American Control Conference.*, vol. 5, Anchorage, AK, May 2002, pp. 3931–3936.

[143] M. Dash and H. Liu, "Feature selection for classification," *Intelligent Data Analysis*, vol. 1, no. 3, pp. 131–156, 1997.

[144] M. Hall and G. Holmes, "Benchmarking attribute selection techniques for discrete class data mining," *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 6, pp. 1437–1447, 2003.

[145] J. Juban, N. Siebert, and G. Kariniotakis, "Probabilistic short-term wind power forecasting for the optimal management of wind generation," in *Power Tech Proceedings, 2007 IEEE Lausane*, Lausane, Zwitserland, July 2007.

[146] J. P. Marques de Sá, *Pattern Recognition: Concepts, Methods ans Applications.* Berlin: Springer Verlag, 2001.

[147] L. M. Costa and G. Kariniotakis, "A stochastic dynamic programming model for optimal use of local energy resources in a market environment," in *Power Tech Proceedings, 2007 IEEE Lausane*, Lausanne, Switzerland, July 2007.

[148] R. Buizza, "Potential Forecast Skill of Ensemble Prediction and Spread and Skill Distributions of the ECMWF Ensemble Prediction System," *Monthly Weather Review*, vol. 125, no. 1, pp. 99–119, 1997. [Online]. Available: http://dx.doi.org/10.1175%2F1520-0493%281997%29125%3C0099%3APFSOEP%3E2.0.CO%3B2

DEVELOPPEMENT DE METHODES POUR LA PREDICTION DE LA PRODUCTION EOLIENNE REGIONALE

Résumé

L'intégration à grande échelle de l'énergie éolienne dans les réseaux électriques peut poser des problèmes aux opérateurs de ces réseaux car, contrairement aux moyens de production conventionnels, la production éolienne est variable et non contrôlable. Pour réduire l'impact de certains de ces problèmes, les gestionnaires de réseaux expriment le besoin de prévisions à court terme (de 48 à 120 heures) de la production agrégée des parcs éoliens situés dans une région définie.
Le but de la thèse est de développer un cadre d'analyse et des outils permettant de faciliter la mise en place de modèles de prévision de la production éolienne régionale.
La thèse présente tout d'abord un cadre d'analyse permettant de caractériser la production éolienne régionale. Par ce biais, les propriétés saillantes de la production régionale, qui doivent être prises en compte lors de la conception d'un modèle de prévision régionale, sont identifiées.
Le problème de la prévision régionale est ensuite abordé comme un problème d'apprentissage statistique. Nous définissons trois approches de modélisation générique permettant la combinaison de sous-modèles. L'influence de ces approches sur la précision des prévisions est étudiée ainsi que celle du choix des sous-modèles. Pour permettre la comparaison de sous-modèles, nous introduisons un modèle de prévision éolienne dont la performance est comparable aux modèles de l'état de l'art.
Finalement, nous examinons l'impact sur la précision de prévision qu'a le choix des variables explicatives et nous proposons des règles générales de sélection dans le cadre de la prévision éolienne régionale. Pour faciliter le processus de modélisation, des méthodes de sélection automatique sont étudiées. Deux méthodes (une méthode filtre et une méthode wrapper) qui exploitent les caractéristiques propres au problème sont proposées. Nous montrons que ces méthodes sont plus performantes qu'une méthode générique de l'état de l'art.

Mots clés : énergie éolienne, énergies renouvelables, prédiction de la production éolienne, intelligence artificielle, sélection de variables explicatives, mathématiques appliquées

DEVELOPMENT OF METHODS FOR REGIONAL WIND POWER FORECASTING

Abstract

The large-scale integration of wind power can be a challenge for power system operators because, unlike conventional power sources, wind power is variable and non-dispatchable. To alleviate some of the problems posed by large-scale wind power integration, power system operators express the need for short-term (48 to 120 hours ahead) forecasts of the aggregated output of all wind farms within a specified geographical region.
The aim of the thesis is to develop a framework and tools to help in the implementation of statistical regional wind power forecasting models.
We first propose a framework for the characterization of the regional wind power. In this way, salient aspects of the regional wind power forecasting problem that must be taken into account when designing a regional forecasting model are identified.
We then examine the regional forecasting problem from a statistical learning perspective. We define three generic approaches that can be used to combine sub-models to build regional models. The influence of these approaches on forecast accuracy is examined, as well as that of the choice of sub-models. The comparison of sub-models is made possible by the introduction of a novel forecasting model whose performance is shown to be comparable to that of other state-of-the-art models.
Finally, we examine the impact of explanatory variable selection on forecast accuracy and derive general guidelines applicable in the frame of regional wind power forecasting. To ease modelling, automatic selection techniques are investigated. Two variable selection methods (a filter and a wrapper method) that exploit problem-specific characteristics are proposed. These methods are shown to compare very favourably to a generic state-of-the-art method.

Key wind power, renewable energies, wind power forecasting, artificial intelligence, explanatory variable selection, applied mathematics

Laboratoire d'accueil :	Centre Energétique et Procédés - Ecole des Mines de Paris Sophia Antipolis : Rue Claude Daunesse - B.P. 207 - F-06904 Sophia Antipolis Cedex
Thèse présentée par :	SIEBERT, Nils le : 6 mars 2008
Discipline :	"Energétique" - Ecole des Mines de Paris